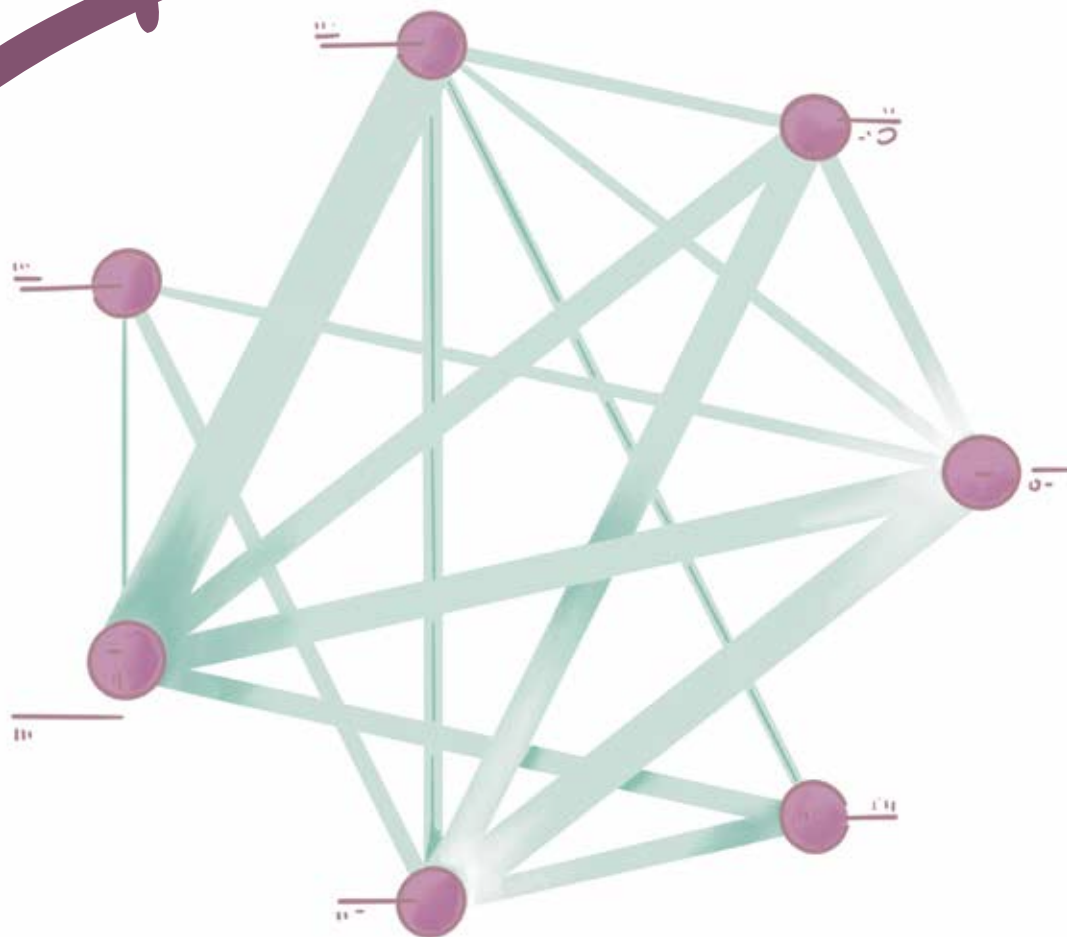


Epidemiología y Estadística



Coordinador

Antonio Martín Conejero

Autores

Marcos Alonso García
Antonio Martín Conejero

Víctor Quirós González

Leyre Rodríguez Leal
Toni Soriano Colomé

Director de la obra

Fernando de Teresa Galván

Índice

Epidemiología

01. Conceptos de epidemiología.		04. Tipos de estudios epidemiológicos	17
Estudio de un test	1	4.1. Tipos de estudio	18
1.1. Conceptos de epidemiología	2	4.2. Estudios descriptivos	18
1.2. Estudio de un test	3	4.3. Estudios analíticos	19
1.3. Relación entre la prevalencia y los valores predictivos	4		
1.4. Aceptabilidad de un método diagnóstico	5	05. El ensayo clínico	23
1.5. Cribado	5	5.1. Introducción	24
1.6. Errores en las pruebas de diagnóstico y cribado	5	5.2. Etapas en la realización de un ensayo clínico	25
1.7. Evaluación de la concordancia	6	5.3. Tipos de ensayos clínicos	27
		5.4. Recomendaciones para la publicación de ensayos clínicos	29
02. Formas de medida	7	5.5. Agencia Española de Medicamentos y Productos Sanitarios	29
2.1. Medidas de la frecuencia de la enfermedad	8		
2.2. Medidas de asociación	9	06. El metanálisis	31
2.3. Medidas de impacto o del efecto	10	6.1. Introducción	32
		6.2. Etapas del metanálisis	32
03. Validez y fiabilidad de los estudios epidemiológicos	13		
3.1. Tipos de error	14	07. Medicina basada en la evidencia	35
3.2. Validez y fiabilidad	15	7.1. Niveles de calidad de la evidencia científica	36
3.3. Validez de un test diagnóstico	15	7.2. El método científico	38

Estadística

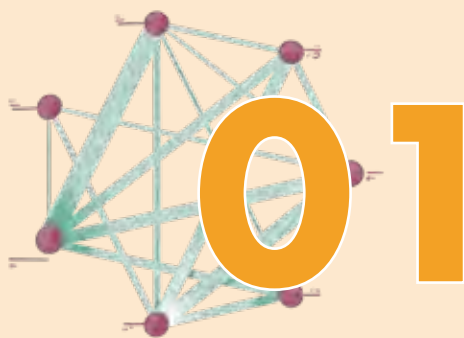
08. Estadística. Generalidades	41	10. El contraste de hipótesis	51
8.1. Tipos de estadística	42	10.1. Hipótesis nula, hipótesis alternativa y grado de significación estadística	52
8.2. Población y muestra	42	10.2. Errores α y β	52
8.3. Variables	42	10.3. Pruebas de significación estadística	52
8.4. Representación gráfica	42	10.4. Pruebas a una cola y a dos colas	54
8.5. Síntesis de los datos	43	10.5. Análisis multivariante	54
8.6. Asimetría y curtosis	44	10.6. Regresión lineal	54
		10.7. Regresión logística	55
09. Principales leyes de distribución de las variables aleatorias	47	10.8. Regresión de COX	56
9.1. Distribución binomial	48		
9.2. Distribución de Poisson	48	11. El tamaño muestral	57
9.3. Distribución normal o de Gauss	48	11.1. Las muestras	58
9.4. Estimación de medias	48		



Planificación y gestión

12. La salud y los sistemas sanitarios59	15. La calidad asistencial y la seguridad del paciente71
12.1. Introducción60	15.1. Introducción72
12.2. Los sistemas sanitarios60	15.2. Componentes de la calidad asistencial72
12.3. El Sistema Nacional de Salud español61	15.3. Monitorización, evaluación, normalización y certificación de la calidad. Modelos de gestión de la calidad72
12.4. La planificación sanitaria61	15.4. La seguridad del paciente73
13. La atención sanitaria63	
13.1. Atención primaria64	
13.2. Atención hospitalaria o especializada64	
14. La evaluación económica67	
14.1. Eficacia, efectividad y eficiencia68	
14.2. Tipos de costes68	
14.3. Evaluación de la eficiencia: análisis de costes69	

Bibliografía75



Conceptos de epidemiología. Estudio de un test

Orientación MIR

Este es un tema importante y sencillo. Hay que conocer los diferentes conceptos, así como resolver tablas que incluyan los conceptos que se resumen en las Ideas clave.

Preguntas MIR

- MIR 23-24, 43; MIR 23-24, 46
- MIR 22-23, 45; MIR 22-23, 189
- MIR 21-22, 44
- MIR 20-21, 50; MIR 20-21, 52; MIR 20-21, 185-IF
- MIR 19-20, 32; MIR 19-20, 33; MIR 19-20, 154; MIR 19-20, 177
- MIR 18-19, 117; MIR 18-19, 129; MIR 18-19, 130; MIR 18-19, 131; MIR 18-19, 206
- MIR 17-18, 194; MIR 17-18, 222
- MIR 16-17, 129
- MIR 15-16, 205; MIR 15-16, 206
- MIR 14-15, 194; MIR 14-15, 195; MIR 14-15, 235
- MIR 13-14, 206; MIR 13-14, 207

Conceptos clave

- ⦿ La demostración experimental supone la mejor evidencia de causalidad.
- ⦿ El principal inconveniente de la demostración experimental son los problemas éticos.
- ⦿ La sensibilidad y la especificidad se refieren respectivamente a la capacidad del test para detectar correctamente a los individuos enfermos y a los sanos. La sensibilidad y la especificidad no dependen de la prevalencia.
- ⦿ Los valores predictivos dependen de la prevalencia. El VPP se refiere a la probabilidad de ser enfermo cuando el test es positivo, y el VPN a la probabilidad de estar sano cuando el test ha sido negativo.
- ⦿ Los test sensibles sirven para descartar enfermedad. Los específicos, para confirmar enfermedad.
- ⦿ Las curvas COR representan la relación recíproca entre sensibilidad y especificidad. Cuanto mayor es el área bajo la curva COR, mejor será el test.
- ⦿ La prevalencia va a determinar el tipo de test más útil en una comunidad. Si la prevalencia es alta, se realizará un test sensible; si es baja, uno específico.

1.1. Conceptos de epidemiología

Definición de epidemiología

Etimológicamente, la palabra *epidemiología* procede del griego y significa tratado sobre el pueblo ('epi': sobre, 'demo': pueblo, 'logos': tratado).

Se puede definir la **epidemiología** como la ciencia que estudia la distribución y los determinantes del fenómeno salud-enfermedad en las poblaciones humanas. La comunidad sustituye al individuo a la hora de buscar la respuesta a ciertas preguntas sobre la etiología y la prevención de la enfermedad, así como acerca de los recursos necesarios para los cuidados de salud de esa población.

La epidemiología se ocupa de dos aspectos fundamentales:

- Estudiar la distribución de las enfermedades en relación con las variables lugar, tiempo y persona. Es lo que se denomina epidemiología descriptiva.
- Buscar los factores que determinan la distribución encontrada e identificar las asociaciones causales. Es lo que hace la llamada **epidemiología analítica**.

Las principales posibilidades de aplicación de la epidemiología son:

- Establecer el diagnóstico de salud de una comunidad.
- Conocer los factores causales de una enfermedad y las probabilidades de enfermar.
- Evaluar la eficacia/efectividad/eficiencia de los procedimientos preventivos y terapéuticos de los servicios de salud.

Los pasos fundamentales dentro del método epidemiológico son:

1. Observar un fenómeno.
2. Elaborar una hipótesis.
3. Probar la hipótesis.
4. Emitir un informe o ley.

Indicadores de riesgo. Causas de enfermedad

Conceptos

Tenemos que tener claros los siguientes conceptos:

- **Riesgo**. Es la probabilidad de desarrollar una enfermedad por estar expuesto a ciertos factores.
- **Causa**. Es la condición que, sola o acompañada de otras condiciones, inicia o permite iniciar una secuencia de acontecimientos que producen un determinado efecto.
- **Factor de riesgo**. Variable endógena o exógena al individuo, controlable, que precede al comienzo de la enfermedad y que está asociada a un incremento de la probabilidad de aparición de la misma, y a la que se puede atribuir "responsabilidad" en la enfermedad (p. ej., tabaco como factor de riesgo de cáncer de pulmón).
- **Marcador de riesgo**. Variable no controlable, endógena al individuo (constitucional), y cuya presencia anuncia a los individuos particularmente vulnerables (p. ej., sexo femenino como marcador de riesgo de cáncer de mama).

- **Indicador de riesgo**. Variable sin relación causal con el problema, pero cuya presencia alerta precozmente sobre el padecimiento (p. ej., manchas de Koplik como signo precursor de la aparición del sarampión).

Criterios epidemiológicos de causalidad

La existencia de asociación estadística no es sinónimo de causalidad, sino que para que se dé, deben cumplirse los denominados criterios de causalidad, que son los criterios de Bradford-Hill ▶ **MIR 18-19, 117** :

- **Fuerza de asociación**. Es un criterio muy importante, consistente en cuántas veces más riesgo de enfermedad tienen las personas expuestas al factor estudiado que las no expuestas.
- **Efecto dosis-respuesta**. Es decir, que la asociación tenga coherencia interna, que al incrementarse la presencia del factor de riesgo aumente la de la enfermedad.
- **Secuencia temporal**. La exposición al presunto factor de riesgo debe preceder al efecto ▶ **MIR 19-20, 32** .
- **Coherencia externa o consistencia de los resultados**. Los resultados deben ser concordantes con los obtenidos en otros estudios (reproductibilidad).
- **Ausencia de distorsiones metodológicas en el estudio**. Se deben a sesgos.
- **Ausencia de explicaciones alternativas**. Esto es, ausencia de otras hipótesis que expliquen los resultados.
- **Plausibilidad biológica**. Que los resultados sean compatibles con el marco de los conocimientos científicos previos.
- **Efecto de la cesación o reversibilidad**. Es decir, que exista una disminución del riesgo tras la reducción de la exposición al presunto factor de riesgo.
- **Demostración experimental**. Es la prueba causal por excelencia, pero su realización plantea, en ocasiones, problemas éticos.

Modelos causales

Son los siguientes:

- **Determinista o unicausal**. Este modelo propugna que siempre que se presente la causa (causa suficiente) y solo cuando se presente la causa (causa necesaria), ocurre el efecto. Hay especificidad de causa (el efecto tiene una única causa) y también de efecto (cada causa produce un único efecto).
- **Multicausal**. Postula que hay pluralidad de causas y multiplicidad de efectos, formando redes causales, de forma que diferentes causas producen el mismo efecto y una única causa ocasiona varios efectos.
- **Determinista modificado (Rothman)**. En este modelo se distinguen tres tipos de causas:
 - **Causa suficiente**. Es aquella que inevitablemente produce el efecto cuando se presenta.
 - **Causa complementaria o contribuyente**. Son aquellas causas que forman parte de una causa suficiente.
 - **Causa necesaria**. Es la que debe estar presente inevitablemente para que se produzca el efecto. Por tanto, ha de formar parte de todas las causas suficientes de una enfermedad.

Niveles de prevención

Los niveles de prevención se clasifican en:

- **Prevención primaria**. Evitar la adquisición de la enfermedad (vacunación, eliminación y control de riesgo ambiental, educación sanitaria...).



- **Prevención secundaria.** Detección precoz de la enfermedad en los que el establecimiento de medidas adecuadas puede impedir la progresión (cribado...) ► MIR 22-23, 189.
- **Prevención terciaria.** Tratamiento y rehabilitación de la enfermedad para enlentecer su progresión y prevenir complicaciones, intentando mejorar la calidad de vida de los pacientes.
- **Prevención cuaternaria.** Medidas adoptadas para identificar a los pacientes en riesgo de un exceso de medicación, protección de una nueva invasión médica evitando intervenciones médicas innecesarias y sugerirle únicamente intervenciones "éticamente aceptables" ► MIR 19-20, 154; MIR 18-19, 206.

1.2. Estudio de un test

La actividad epidemiológica estudia la frecuencia de enfermedad. Sin embargo, todas sus medidas son realmente de la frecuencia de diagnósticos de enfermedad, de ahí la importancia de conocer la auténtica correspondencia entre el diagnóstico y la realidad patológica.

Muy pocas pruebas diagnósticas, quizá ninguna, identifican con certeza si el paciente padece o no la enfermedad.

La eficacia de una prueba diagnóstica depende de su capacidad para detectar correctamente la presencia o la ausencia de la enfermedad que se estudia, lo que se expresa matemáticamente mediante los cuatro índices: sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. Los dos primeros índices (sensibilidad y especificidad) marcan lo que se denomina **validez interna del test** y los dos segundos, la **validez externa**, esto es, la que se obtiene al aplicar esa prueba en un entorno poblacional determinado.

Los citados índices se obtienen a partir del análisis de una serie de pacientes a los que se les realiza una prueba diagnóstica (prueba problema), comparándose los resultados con los de una prueba de superior rendimiento diagnóstico (prueba de referencia). Los resultados obtenidos se expresan de la siguiente manera (Tabla 1.1) ► MIR 20-21, 185-IF :

	ENFERMO	SANO	TOTAL
Positivo	VP (a)	FP (b)	Total de positivos (a + b)
Negativo	FN (c)	VN (d)	Total de negativos (c + d)
Total	Total de enfermos (a + c)	Total de sanos (b + d)	Total de individuos (a + b + c + d)

Verdaderos positivos (VP): resultados positivos en sujetos enfermos.
 Verdaderos negativos (VN): resultados negativos en sujetos sanos.
 Falsos positivos (FP): resultados positivos en sujetos sanos.
 Falsos negativos (FN): resultados negativos en sujetos enfermos.

Tabla 1.1. Resultado de la prueba diagnóstica evaluada.

- **Sensibilidad (S).** Se define como la probabilidad de que un individuo enfermo tenga un test positivo. La sensibilidad indica la proporción del total de enfermos que el test es capaz de detectar ► MIR 14-15, 235.

$$S = \frac{\text{Individuos enfermos con test (+)}}{\text{Total enfermos}} = \frac{VP}{VP + FN} = \frac{a}{a + c}$$

- **Especificidad (E).** Es la probabilidad de que un individuo sano tenga un test negativo. Indica la proporción de individuos sanos confirmados como tales por el resultado negativo del test ► MIR 21-22, 44; MIR 13-14, 207.

$$E = \frac{\text{Individuos sanos con test (-)}}{\text{Total sanos}} = \frac{VN}{VN + FP} = \frac{d}{b + d}$$

- **Tasa de falsos negativos (TFN).** Es la probabilidad de que un individuo, estando enfermo, sea clasificado como sano.

$$TFN = \frac{\text{Individuos enfermos con test (-)}}{\text{Total enfermos}} = \frac{FN}{VP + FN} = \frac{c}{a + c}$$

- **Tasa de falsos positivos (TFP).** Mide la probabilidad de que a un individuo sano se le clasifique como enfermo.

$$TFP = \frac{\text{Individuos sanos con test (+)}}{\text{Total sanos}} = \frac{FP}{VN + FP} = \frac{b}{b + d}$$

- **Razón de probabilidad (RP) o verosimilitud positiva.** Compara la probabilidad de que un individuo enfermo presente un resultado positivo frente a la probabilidad de que el resultado positivo se presente en un individuo sano ► MIR 20-21, 52; MIR 19-20, 177; MIR 15-16, 205; MIR 12-13, 196. Es un parámetro muy útil porque dice cuánto es más probable que aparecido un test positivo, dicho resultado provenga de un enfermo y no de un sano. Interesa, por tanto, que sea siempre mayor de 1, lo que indicará que la probabilidad de que venga el resultado positivo del enfermo es mayor que del sano, y cuanto mayor, mejor, porque aumenta dicha probabilidad a favor del enfermo.

$$RP (+) = \frac{\text{Probabilidad test (+) en enfermos}}{\text{Probabilidad test (+) en sanos}} = \frac{VP/\text{enfermos}}{FP/\text{sanos}}$$

$$RP (+) = \frac{S}{1 - E}$$

- **Razón de probabilidad o verosimilitud negativa.** Se relaciona la probabilidad de que un individuo enfermo presente un resultado negativo con la probabilidad de que el resultado negativo se dé en un individuo sano. El parámetro hace referencia a la probabilidad de que aparecido un test negativo, este venga de un enfermo con respecto a la probabilidad de que sea sano (Tabla 1.2). Interesa, por tanto, que este parámetro sea siempre menor de 1 (aparecido un negativo, será más probable que venga de un sano), y cuanto más pequeño, mejor.

$$RP (-) = \frac{\text{Probabilidad test (-) en enfermos}}{\text{Probabilidad test (-) en sanos}} = \frac{FN/\text{enfermos}}{VN/\text{sanos}}$$

$$RP (-) = \frac{1 - S}{E}$$

Valor de CP+	1-2	2-5	5-10	> 10
Valor de CP-	1-0,5	0,5-0,2	0,2-0,1	< 0,1
Interpretación	Variación insignificante	Pequeños cambios	Cambios moderados	Grandes cambios concluyentes

Tabla 1.2. Variación desde una probabilidad preprueba a una probabilidad posprueba.

- **Valor predictivo positivo.** Se trata de la proporción de verdaderos positivos entre aquellos que han sido identificados como positivos en una prueba de test **MIR 18-19, 129; MIR 16-17, 129; MIR 14-15, 195**.

$$VPP = \frac{VP}{VP + FP} = \frac{a}{a + b}$$

- **Valor predictivo negativo.** Se trata de la proporción de verdaderos negativos entre aquellos que han sido identificados como negativos en un test **MIR 22-23, 45**.

$$VPN = \frac{VN}{VN + FN} = \frac{d}{c + d}$$

- **Valor global (eficacia) del test.** Indica la proporción de resultados válidos entre el conjunto de resultados.

$$VGT = \frac{VP + VN}{VP + VN + FP + FN} = \frac{a + d}{\text{Total}}$$

- Sensibilidad + tasa de falsos negativos = 100%.
- Especificidad + tasa de falsos positivos = 100%.

Recuerda

- La sensibilidad y la especificidad clasifican a los sujetos en función del estatus de enfermedad. Los valores predictivos lo hacen en función del resultado del test.

El resultado de un test puede ser continuo (p. ej., niveles de glucemia en mg/dL) y entonces hay que decidir cuál se considerará como resultado positivo del mismo, por lo que es necesario elegir un punto de corte (**Figura 1.1**).

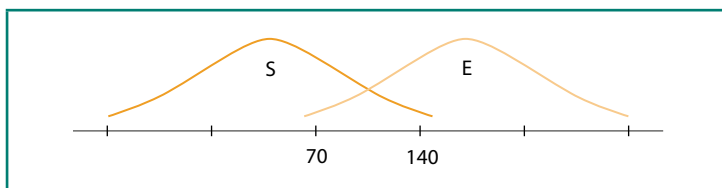


Figura 1.1. Variación del punto de corte para determinar la sensibilidad y especificidad.

El punto de corte escogido determinará la sensibilidad y la especificidad de la prueba (si se elige 70, la prueba será muy sensible y poco específica; si se selecciona 140, será poco sensible y muy específica). Los test diagnósticos cuantitativos cambian su sensibilidad y especificidad de forma inversa y siempre dependiendo de cómo varía el punto de corte del test que se está evaluando **MIR 13-14, 206**.

Para determinar el punto de corte, se pueden utilizar las curvas de características operativas para el receptor (COR). Se seleccionan varios puntos de corte y se estima la sensibilidad y la especificidad para cada uno de ellos **MIR 23-24, 46**. Posteriormente, se representa gráficamente la sensibilidad en función de (1 - E). La prueba ideal se sitúa en el ángulo superior izquierdo (S y E = 1). En general, para hacer la elección del mejor punto de corte se suele emplear lo que se denomina *área bajo la curva* (COR, o ROC en su terminología inglesa); a mayor área bajo la curva, mejor capacidad de discriminación de esa prueba.

Una prueba sin ningún valor sigue la diagonal que va del ángulo inferior izquierdo al ángulo superior derecho (cada incremento de la sensibilidad se asocia a una pérdida de igual magnitud de la especificidad). De esta forma, la sensibilidad y la especificidad son valores interdependientes, de manera que si aumenta la sensibilidad, disminuye la especificidad, y viceversa (**Figura 1.2**).

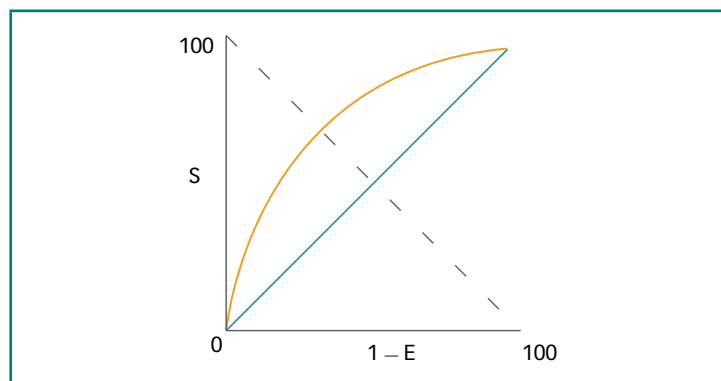


Figura 1.2. Área bajo la curva (COR).

Si se adoptan criterios de diagnóstico muy estrictos, decrece la sensibilidad (hay menos enfermos que cumplen estos criterios), y paralelamente se incrementa la especificidad (pocos sanos cumplen los criterios citados).

1.3. Relación entre prevalencia y valores predictivos

Los valores predictivos de un test son variables, dependen de la prevalencia de la enfermedad en la población. La sensibilidad y la especificidad son características propias del test y no se modifican con cambios en la prevalencia **MIR 20-21, 50; MIR 15-16, 206**. Si la prevalencia de la enfermedad aumenta, se incrementa el valor predictivo positivo, mientras que disminuye el valor predictivo negativo.

Esto es fácil de entender, por ejemplo, en el caso del VPP; si en una población la prevalencia de una enfermedad es elevada, tendrá mayor probabilidad de hacer test a individuos realmente enfermos, lo que disminuirá la tasa de falsos positivos y, por ello, aumentará el VPP.

Por el contrario, si la prevalencia de una enfermedad es muy pequeña, se puede tener una alta probabilidad de realizar test que den resultados falsos positivos, porque en la mayor parte el test se estará aplicando sobre individuos sanos, incluso con cifras de buena sensibilidad del test, el VPP disminuye, en ocasiones notablemente. Por eso es tan importante conocer no solo los valores de sensibilidad y especificidad sino el ámbito en el que se aplican para entender los conceptos de VPP y VPN. Por el contrario, si la prevalencia de la enfermedad decrece, aumenta el valor predictivo negativo y se reduce el valor predictivo positivo.



Recuerda

- ↑ Prevalencia: ↑ VPP y ↓ VPN.
- ↓ Prevalencia: ↓ VPP y ↑ VPN.

1.4. Aceptabilidad de un método diagnóstico

No existe un parámetro guía útil para todas las situaciones. La aceptabilidad de un test depende de la patología estudiada y de las condiciones reales en el medio y en la colectividad.

Si lo que interesa es detectar el mayor número posible de enfermos, se debe usar un test con alta sensibilidad. Así se escaparán pocos enfermos, aunque con un coste de obtención de bastantes falsos positivos.

Se elegirá un test sensible cuando:

- La enfermedad sea grave y no pueda pasar desapercibida.
- La enfermedad sea tratable.
- Los resultados falsamente positivos no supongan un trauma psicológico en los individuos examinados.

Recuerda

- Como un test sensible genera muchos falsos positivos, la rentabilidad del mismo es que su resultado sea negativo.

Si lo que se quiere es asegurar el diagnóstico, se ha de emplear un test cuya especificidad sea máxima. Se utilizará un test lo más específico posible cuando ► MIR 18-19, 131 :

- La enfermedad sea importante, pero difícil de curar o incurable.
- Los resultados falsamente positivos puedan suponer un trauma psicológico para el individuo examinado.
- El tratamiento de los falsos positivos pudiera tener graves consecuencias.

1.5. Cribado

El **cribado** o *screening* es una estrategia de detección precoz de la enfermedad. Lo ideal es aplicar primero un test muy sensible (que detecta todos los casos posibles de enfermedad; por tanto, se obtienen muchos FP y pocos FN) y en una segunda fase, aplicar un test muy específico (con el que se confirma el diagnóstico de esos posibles enfermos; se obtienen muy pocos FP) ► MIR 17-18, 222 .

En la puesta en marcha de un programa de cribado deben tenerse en cuenta los criterios dependientes de la enfermedad, del test y de la población diana que se detallan a continuación. Sin embargo, no todos los programas de cribado necesariamente pueden resultar beneficiosos.

El sobrediagnóstico es el fenómeno que ocurre cuando las personas son diagnosticadas de enfermedades que nunca les causarían síntomas ni les

acortarían la vida, pero en cambio, la alarma generada por el hallazgo les puede ocasionar consecuencias indeseables ► MIR 17-18, 194 .

Los cribados pueden estar en función de:

- **Criterios dependientes de la enfermedad:**
 - La enfermedad debe ser común y grave.
 - Ha de conocerse la historia natural de la enfermedad.
 - El tratamiento, en el estado presintomático, tiene que reducir la morbilidad en mayor medida que el tratamiento después de la aparición de los síntomas.
- **Criterios dependientes del test** ► MIR 23-24, 43 :
 - De fácil aplicación.
 - De coste razonable dentro del presupuesto de salud.
 - Inocuos para la población.
 - Confiabilidad o capacidad de repetición.
 - Validez. Se refiere a la capacidad del test de medir lo que realmente se desea medir. El concepto de validez incluye los de sensibilidad, especificidad y valor predictivo; ya se han mencionado previamente los conceptos de validez interna y externa.
- **Criterios dependientes de la población diana:**
 - El riesgo de ser afectado por la enfermedad debe ser alto.
 - La información demográfica tiene que estar disponible en la comunidad.
 - La comunidad ha de sentir la necesidad de programas de salud pública.

■ Clasificación de los cribados

Los cribados se clasifican en:

- **Cribado simple** (para una enfermedad) o **múltiple** (para varias enfermedades). Lo habitual es que el cribado se dedique a una sola enfermedad.
- **Cribado monofásico** (de una etapa) o **multifásico** (de varias etapas). Las pruebas que normalmente se utilizan suelen ser muy sensibles pero no tan específicas, por lo que los positivos dados por la prueba de cribado han de confirmarse posteriormente con otras. La mayoría de los cribados son, en consecuencia, multifásicos.
- **Cribado ocasional** (cuando un individuo acude por otro motivo a un servicio sanitario) o **sistemático** (dirigido al conjunto de la población).
- **Cribado no selectivo** (dirigido a toda la población) o **selectivo** (dirigido a grupos de especial riesgo elevado).
- **Cribado precoz** (búsqueda de la enfermedad en su estadio inicial) o **tardío** (búsqueda de casos "olvidados" en la población). La mayor parte de los cribados intentan ser precoces ► MIR 14-15, 194 .

1.6. Errores en pruebas de diagnóstico y cribado

En ocasiones, algunos programas de cribado y pruebas diagnósticas pueden implicar que cometamos ciertos sesgos, entre los que se encuentran ► MIR 19-20, 33 :

- **Sesgo por exclusión de indeterminados.** Este sesgo se produce cuando solo son incluidos en el estudio aquellos pacientes que tienen resultados en la prueba a validar claramente positivos o claramente negativos. De esta forma la concordancia con el *gold standard* es mayor que si no se excluyera ningún sujeto.
- **Sesgo de duración (*length bias*).** Consiste en que en las pruebas de cribado se suelen detectar las formas más benignas de la enfermedad dado

que suelen tener un curso más larvado. Se escapan sin embargo las formas más agresivas con menor fase asintomática por lo que mayor detección de casos no necesariamente se acompaña de menor mortalidad.

- **Sesgo de adelanto diagnóstico (*lead time bias*).** En este caso la prueba de cribado detecta antes la enfermedad, pero no cambia su pronóstico. Por tanto, el paciente no vive más, si no que vive más tiempo diagnosticado.
- **Sesgo de revisión.** Se produce cuando el que tiene que interpretar el resultado de la prueba a validar conoce el resultado del *gold standard*. Esto tiende a variar el resultado de la prueba que se quiere validar.
- **Reproducibilidad.** El resultado de la validación de una prueba es igual, aunque se repita en diferentes situaciones.
- **Sesgo de verificación.** Consiste en que se puede clasificar de diferente manera al sujeto que da una prueba inicial negativa, porque no se hace en este grupo el mismo tipo de pruebas que si sale positiva la prueba inicial ► **MIR 18-19, 130**.
- **Sobrediagnóstico.** Es el fenómeno que ocurre cuando las personas son diagnosticadas de enfermedades que nunca les causarían síntomas ni les acortaría la vida, pero en cambio, la alarma generada por el hallazgo les puede ocasionar consecuencias indeseables ► **MIR 17-18, 194**.

1.7. Evaluación de la concordancia

■ Coeficiente de correlación intraclass

Para el caso de variables cuantitativas, es habitual que el análisis de la concordancia se aborde mediante técnicas estadísticas inapropiadas. Con frecuencia, se ha utilizado el cálculo del coeficiente de correlación lineal (r) de Pearson como índice de concordancia, así como la correlación de Spearman. Sin embargo, estas no resultan una medida adecuada del grado de acuerdo entre dos mediciones, ya que si dos instrumentos miden sistemáticamente cantidades diferentes uno del otro, la correlación puede ser perfecta ($r = 1$), a pesar de que la concordancia sea nula. Como se analizará más adelante, el coeficiente de correlación de Pearson no proporciona información sobre el acuerdo observado, y solamente mide la asociación lineal entre dos variables.

Desde el punto de vista matemático, el índice más apropiado para cuantificar la concordancia entre diferentes mediciones de una variable numérica

es el coeficiente de correlación intraclass (CCI). Como toda proporción, los valores del CCI pueden oscilar entre 0 y 1, de modo que la máxima concordancia posible corresponde a un valor de CCI = 1. En este caso, toda la variabilidad observada se explicaría por las diferencias entre sujetos y no por las diferencias entre los métodos de medición o los diversos observadores. Por otro lado, el valor CCI = 0 se obtiene cuando la concordancia observada es igual a la que se esperaría que ocurriera solo por azar.

■ Coeficiente kappa

El **índice kappa** (kappa de Cohen) relaciona el acuerdo que exhiben los observadores, más allá del debido al azar, cuando la variable de análisis es cualitativa. Se calcula la diferencia entre la proporción de acuerdo observado y la proporción de acuerdo esperado por azar; si esta es igual a 0, entonces el grado de acuerdo que se ha observado puede atribuirse enteramente al azar; si la diferencia es positiva, ello indica que el grado de acuerdo es mayor que el que cabría esperar si solo estuviera operando el azar y viceversa; en el caso (ciertamente improbable) en que la diferencia fuera negativa, entonces los datos estarían exhibiendo menos acuerdo que el que se espera solo por concepto de azar.

Una situación especial es aquella en la que las variables cualitativas son ordinales. En este caso es apropiado el empleo del índice kappa ponderado. Supóngase que las categorías de clasificación son, por ejemplo, "sano", "posiblemente enfermo" y "claramente enfermo".

A la hora de valorar el grado de discrepancia entre dos observadores, no es lo mismo que uno clasifique a un sujeto como "posiblemente enfermo" y el otro lo declare "sano", a que uno lo clasifique como "sano" y el otro como "claramente enfermo". La "distancia" entre ambas discrepancias no es la misma. El índice kappa ponderado tiene en cuenta el peso de las diferencias entre las distintas categorías de la variable ordinal.

No obstante, existen **limitaciones del estadístico kappa**:

- El valor de kappa se ve afectado por la prevalencia del rasgo estudiado. Por tanto, es necesario ser cuidadoso a la hora de generalizar los resultados de comparación de observadores en situaciones con prevalencias diferentes; esto quiere decir que kappa es un estadístico descriptivo útil, pero de limitada validez externa.
- Kappa es dependiente del número de categorías. Cuantas más se estén considerando, más difícil será clasificar correctamente los sujetos de observación, lo que habitualmente implica valores de kappa más bajos.

Casos clínicos

La probabilidad preprueba de padecer cáncer de mama en mujeres mayores de 50 años es del 20%, mientras que es del 10% en las que tienen más de 40 años. Señale la afirmación correcta:

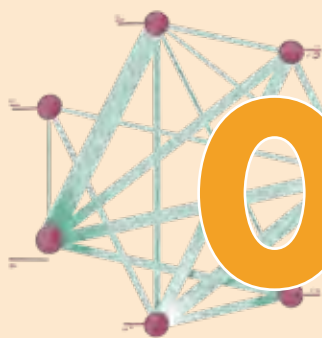
- 1) La sensibilidad de la mamografía es mayor en las mujeres de más de 50 años.
- 2) El VPP será mayor en las mujeres de más de 40 años.
- 3) El VPN será menor en las mujeres de más de 50 años.
- 4) La especificidad varía de forma recíproca a la prevalencia.

RC: 3

Si se hace el diagnóstico de insuficiencia renal crónica con un aclaramiento de creatinina plasmática de 60 en vez de 80, se habrá hecho un test con mayor:

- 1) Sensibilidad.
- 2) Especificidad.
- 3) No se modificará ni sensibilidad ni especificidad.
- 4) VPN.

RC: 2



Formas de medida

Orientación MIR

En las medidas de frecuencia de la enfermedad es importante conocer los conceptos de incidencia, prevalencia y densidad de incidencia.

En las medidas de asociación lo más importante es el concepto de riesgo relativo.

Los conceptos fundamentales de las medidas de impacto o del efecto son: el número de sujetos necesario para tratar (NNT), la reducción absoluta del riesgo (RAR) y la reducción relativa del riesgo (RRR). Es esencial saber cómo se formulan las preguntas que se responden a posteriori con cada uno de los parámetros.

Preguntas MIR

- MIR 23-24, 44
- MIR 21-22, 46
- MIR 20-21, 49; MIR 20-21, 54
- MIR 19-20, 31; MIR 19-20, 34; MIR 19-20, 37
- MIR 18-19, 118; MIR 18-19, 119
- MIR 16-17, 118; MIR 16-17, 134
- MIR 15-16, 191
- MIR 14-15, 131; MIR 14-15, 181; MIR 14-15, 189; MIR 14-15, 198
- MIR 13-14, 195; MIR 13-14, 196; MIR 13-14, 201; MIR 13-14, 205

Conceptos clave

- ⊙ La incidencia es sinónimo de riesgo, porque se refiere a los nuevos casos debidos a una enfermedad. Como incidencia es riesgo, todos los parámetros que incluyen el término riesgo (RR, RAE, RRR, RAR) se calculan utilizando incidencias.
- ⊙ La prevalencia o probabilidad pretest se refiere a la cantidad de enfermos en una población, no solo a los casos nuevos.
- ⊙ Es importante conocer los aspectos de incidencia, prevalencia y densidad de incidencia.
- ⊙ El RR es la medida de asociación de los estudios prospectivos, es decir, de cohortes y de los ensayos clínicos.
- ⊙ Es importante conocer los valores posibles, así como el significado de RR menor, igual o mayor de 1.
- ⊙ Cuando en una pregunta, además del RR, se proporcione un intervalo de confianza, hay que fijarse en este último. Si el intervalo incluye el 1, entonces se dice que el RR no ha sido significativo.
- ⊙ La OR es la medida de asociación de los diseños retrospectivos o de casos y controles. Su mayor utilidad es cuando la enfermedad de estudio es poco prevalente.
- ⊙ El NNT siempre se pregunta de la misma forma: ¿a cuántos pacientes se debería tratar para conseguir un evento?
- ⊙ El NNT es tanto mejor cuanto menor sea.
- ⊙ La RRR y la RAR sirven para cuantificar la relevancia clínica de una medida terapéutica. Su cálculo se realiza mediante una resta de incidencias.
- ⊙ El RA y la FAE cuantifican cuánto de la enfermedad se debe a un factor de riesgo, en el grupo de sujetos expuestos a dicho factor.

2.1. Medidas de frecuencia de la enfermedad

■ Formas básicas de medida

Son las siguientes:

- **Número.** Es un valor absoluto. Al no conocer el denominador no se puede saber la importancia relativa del problema.
- **Razón.** Es el cociente en el que el numerador no está incluido en el denominador, es decir, cociente entre dos entidades que poseen caracteres distintos. Sus valores oscilan entre 0 e infinito.
 - Fórmula: $R = a / b$
 - Ejemplo: razón de hombres / mujeres.
- **Proporción.** Es el cociente en el cual el numerador está incluido en el denominador. Sus valores oscilan entre 0 y 1. Es adimensional.
 - Fórmula: $P = a / a + b$
 - Ejemplo: proporción de aprobados en un examen.
- **Tasa.** Mide la ocurrencia de un suceso durante un período de tiempo en una población determinada. Es un cociente en el que el numerador está incluido en el denominador, pero donde, a diferencia de la proporción, el tiempo también está incluido en este. Tiene, por tanto, dimensión temporal.
 - Fórmula: $T = a / \text{persona} \times \text{tiempo}$.
 - Ejemplo: 300 cánceres de pulmón / 100.000 personas-año.

■ Medidas de frecuencia de la enfermedad

Las medidas de frecuencia de la enfermedad se llevan a cabo con los siguientes indicadores:

- **Prevalencia (P).** Es el número total de casos entre el total de la población. No hace distinción entre casos antiguos y casos nuevos y describe la situación en un momento determinado del tiempo. Su valor es más alto a mayor incidencia y duración de la enfermedad **► MIR 14-15, 131**.

$$P = \frac{\text{N.º total de casos de enfermedad}}{\text{Población en ese momento}}$$

- **Incidencia acumulada (IA).** Se define como el número de casos nuevos de una enfermedad que se producen en un período de tiempo. Es la probabilidad de que un individuo desarrolle la enfermedad en ese período **► MIR 19-20, 31**.

$$IA = \frac{\text{N.º casos nuevos en un período de tiempo}}{\text{Población en riesgo al inicio de ese período}}$$

- **Densidad de incidencia (DI).** Se define como el número de casos nuevos de la enfermedad que se producen en un período de tiempo, teniendo en cuenta la información que aporta el sumatorio de riesgo de cada uno de los individuos **► MIR 19-20, 37**.

$$DI = \frac{\text{N.º casos nuevos}}{\text{Sumatorio de personas} \times \text{tiempo}}$$

Personas × tiempo. Es una unidad de medida de la población en riesgo. Se obtiene sumando la cantidad de tiempo que cada individuo ha estado expuesto al factor de estudio; p. ej., durante el seguimiento de 10 sujetos expuestos a un factor de riesgo, se ha observado:

SUJETO	APARECE LA ENFERMEDAD	TIEMPO DE SEGUIMIENTO
1	No	5
2	No	5
3	Sí	4
4	No	3
5	Sí	5
6	Sí	1
7	Sí	3
8	No	5
9	No	5
10	No	5

Se puede comprobar que 4 sujetos han desarrollado la enfermedad, pero que no todos los incluidos en el estudio han tenido el mismo tiempo de seguimiento, por lo que la densidad de incidencia será:

$$DI = \frac{4 \text{ casos nuevos}}{(6 \times 5) + (1 \times 4) + (2 \times 3) + (1 \times 1)}$$

La forma de calcular la DI es, por tanto, dividir el número de eventos nuevos durante el período de estudio, entre el tiempo total observado para todas las personas del estudio. Por ejemplo, 10 casos por cada 1.000 personas/año sería el resultado de encontrar 10 casos en 1.000 personas seguidas durante un año o en 500 personas seguidas durante 2 años.

La densidad o tasa de incidencia, por tanto, expresa los casos nuevos de enfermedad que se producen en un período de tiempo que corresponde al sumatorio de los períodos individuales de riesgo **► MIR 13-14, 196**. Este concepto se conoce también con el nombre de **fuerza de la morbilidad**.

■ Medidas de gravedad de la enfermedad

- **Mortalidad.** Se define como proporción entre el número de fallecidos en una población durante un determinado período de tiempo y la población total en ese mismo período.

$$\text{Mortalidad} = \frac{\text{N.º fallecimientos}}{\text{Población total}}$$

- **Letalidad.** Se define como proporción entre el número de fallecidos en una población durante un determinado período de tiempo y el total de enfermos en ese mismo período **► MIR 20-21, 49**.

$$\text{Letalidad} = \frac{\text{N.º fallecimientos}}{\text{Total de enfermos}}$$



MEDIDA DE FRECUENCIA	CÁLCULO	SIGNIFICADO
Prevalencia	$\frac{\text{Casos existentes}}{\text{Población total}}$	Proporción de individuos enfermos en una población en un momento concreto
Incidencia acumulada	$\frac{\text{Casos nuevos observados}}{\text{Población susceptible al comienzo del período}}$	Riesgo individual de enfermarse
Densidad de incidencia	$\frac{\text{Casos nuevos observados}}{\text{Suma de los períodos de riesgo}}$	Velocidad con que determinada enfermedad aparece en una población

Tabla 2.1. Medidas de frecuencia de la enfermedad.

La relación entre estos parámetros es muy importante. Es posible tener una alta mortalidad en enfermedades de baja letalidad, si la incidencia de la enfermedad es muy alta, como por ejemplo en determinados momentos de la pandemia de COVID-19.

Recuerda

- Hay que observar que en las preguntas en las que a partir de datos se pide calcular incidencias, en el denominador habrá que restar los enfermos al inicio del período, mientras que no habrá que restarlos si lo que se pregunta es la prevalencia.

A modo de resumen de lo explicado, la **Tabla 2.1** enumera las medidas de frecuencia de la enfermedad.

2.2. Medidas de asociación

Las **medidas de asociación** sirven para cuantificar cuánto y cómo se asocia un factor de riesgo a una enfermedad, a partir de la frecuencia con la que aparece el evento resultado en un grupo expuesto a un factor de riesgo, frente al grupo que no está expuesto a dicho factor. Es decir, se emplean para indicar si un factor analizado es protector o de riesgo y se utilizan para cuantificar la asociación entre factor de riesgo y evento resultado.

Riesgo relativo

El **riesgo relativo (RR)** es la medida de asociación en los estudios de cohortes. Sirve para medir la “fuerza de la asociación” entre el factor de riesgo y la enfermedad (**Tabla 2.2**). Puede variar entre 0 e infinito ▶ **MIR 23-24, 44; MIR 19-20, 34; MIR 14-15, 181; MIR 13-14, 205**.

	ENFERMOS	SANOS	TOTAL
FR	a	b	a + b
No FR	c	d	c + d
Total	a + c	b + d	a + b + c + d

Tabla 2.2. Cálculo del RR.

El RR responde a la pregunta: ¿cuánto más frecuente es la enfermedad entre los expuestos a un factor de riesgo respecto a los no expuestos?

Su significado varía dependiendo del valor que tome:

- $RR > 1$: factor de riesgo (FR).

- $RR = 1$: es indiferente; la incidencia es igual en expuestos y en no expuestos.
- $RR < 1$: factor de protección.

$$RR = \frac{\text{Incidencia en expuestos}}{\text{Incidencia en no expuestos}}$$

- Incidencia en expuestos: $I_e = a / a + b$
- Incidencia en no expuestos: $I_o = c / c + d$

$$\frac{I_e}{I_o} = \frac{a / a + b}{c / c + d}$$

Odds ratio (razón de desventaja)

La **odds ratio (OR)** es una razón. Constituye la medida básica de los estudios casos-contrroles.

Su significado es idéntico al del RR, aunque no puede calcularse como este, ya que en los estudios casos-contrroles no es posible valorar la incidencia de la enfermedad (**Tabla 2.3**).

	CASOS (ENFERMOS)	CONTROLES (SANOS)	TOTAL
FR	a	c	a + b
No FR	b	d	c + d
Total	a + c	b + d	a + b + c + d

Tabla 2.3. Cálculo de la OR.

$$OR = \frac{a / b}{c / d} = \frac{a \times d}{b \times c}$$

Para que la OR sea un buen estimador del RR, es necesario que los controles sean representativos de la población de la cual han surgido los casos y que la enfermedad tenga una prevalencia baja, es decir, ha de ser inferior al 10% ▶ **MIR 14-15, 189; MIR 13-14, 201**.

$$OR = \frac{\text{Odds de exposición en los casos}}{\text{Odds de exposición en los controles}} = \frac{\frac{\text{Casos expuestos}}{\text{Casos no expuestos}}}{\frac{\text{Controles expuestos}}{\text{Controles no expuestos}}}$$

Razón de prevalencia

La **razón de prevalencia (RP)** es la medida de asociación de los estudios transversales. Su interpretación es similar a la del RR, es decir, el número de veces más que padecen la enfermedad los expuestos frente a los no expuestos.

$$RP = \frac{\text{Enfermos expuestos}}{\text{Enfermos no expuestos}}$$

Interpretación de las medidas de asociación

La lectura de una OR, un RR y una RP es idéntica y basta con aplicar las palabras "veces más" o "por cada 1"; por ejemplo, si un fármaco con respecto al placebo produce un RR de 4, significa que el fármaco es 4 veces mejor que el placebo. Si se tratara de un estudio de casos y controles y se encuentra una OR de 3 entre tabaco y cáncer, se dirá que el grupo de fumadores tiene 3 veces el riesgo de cáncer del no fumador.

Este tipo de resultados se suele acompañar del intervalo de confianza (IC), que es la expresión de lo que la variabilidad debida al azar puede hacerlo oscilar en la población real (expresa los límites que con una cierta seguridad contendrán ese verdadero valor).

Generalmente los IC se calculan con una seguridad del 95%, y solo deja una probabilidad del 5% de que el verdadero valor de la OR no se halle en ese intervalo.

¿Cómo se sabe si es estadísticamente significativo? Los intervalos de confianza no son significativos si se incluyen en valor nulo, que en el caso de las medidas de asociación es el 1, o valor que no permite saber si el factor de estudio es protector o de riesgo ► **MIR 18-19, 118**.

2.3. Medidas de impacto o del efecto

Diferencia de incidencias o riesgo atribuible

El **riesgo atribuible (RA)** o **riesgo atribuible en expuestos (RAE)** o diferencia de incidencias es una medida que informa sobre el exceso de riesgo en los individuos expuestos frente a los no expuestos al factor de riesgo.

Indica la incidencia acumulada en el grupo de expuestos que se debe exclusivamente al factor de riesgo. Representa el descenso en el número de casos nuevos entre los expuestos si se evitara el factor de riesgo ► **MIR 16-17, 134; MIR 14-15, 181**. Se calcula con:

$$RA = I_e - I_o$$

Fracción atribuible en expuestos o fracción etiológica del riesgo

La **fracción atribuible en expuestos (FAE)** o fracción etiológica del riesgo es la proporción de la enfermedad que se debe a la exposición, es decir, la proporción de casos nuevos de enfermedad, entre los expuestos, que se evitaría si se eliminase el factor de riesgo. Por tanto, estas dos medidas son muy importantes a la hora de evaluar, en términos de salud pública, cuál puede ser el impacto con determinadas medidas, calcular los costes y determinar prioridades.

Se calcula de la siguiente manera:

$$FAE = \frac{I_e - I_o}{I_e}$$

Reducción absoluta del riesgo

Cuando se evalúa una medida preventiva, como un factor protector o el resultado de un fármaco frente a placebo en un ensayo clínico, el beneficio de dicha medida se cuantifica como una reducción del número de enfermos nuevos.

Este concepto está representado por la **reducción absoluta del riesgo (RAR)**. Se calcula:

$$RAR = I_o - I_e$$

Reducción relativa del riesgo

De nuevo se emplea para calcular el beneficio de una medida preventiva. En el caso de la **reducción relativa del riesgo (RRR)**, se expresa como porcentaje de casos evitados por haber sido expuestos a dicho factor protector ► **MIR 13-14, 195**.

Se calcula de la siguiente manera:

$$RRR = \frac{I_o - I_e}{I_o}$$

Número de sujetos necesario para tratar

A partir de los parámetros anteriores, puede calcularse el número de sujetos que deberían exponerse al factor preventivo o terapéutico para conseguir una curación, esto es, el número de sujetos necesario para tratar (NNT) ► **MIR 21-22, 46; MIR 20-21, 54; MIR 18-19, 119; MIR 16-17, 118; MIR 15-16, 191**.

Se calcula de la siguiente manera:

$$NNT = \frac{100}{I_o - I_e}$$



■ Número de sujetos necesario para dañar

El **número de sujetos necesario para dañar (NND** o, en inglés, **NNH** [*numbers needed to harm*]), se calcula de forma similar al NNT pero, a diferencia de este último, se emplea para conocer cuántos sujetos se deben tratar para que aparezca un efecto adverso. En este sentido, mientras que el NNT interesa que sea pequeño, el NND interesará que sea grande ► **MIR 14-15, 198**.

■ Significación de las medidas de impacto

Como en las medidas de asociación, en las medidas de impacto los resultados se suelen acompañar del **intervalo de confianza (IC)**. Generalmente, los IC se calculan con una seguridad del 95%. ¿Cómo se sabe si es estadísticamente significativo? Los intervalos de confianza no son significativos si incluyen el valor nulo, que en el caso de las medidas de impacto es el 0.

Casos clínicos

En un estudio sobre el cáncer de pulmón y la exposición al sílice se han obtenido los siguientes resultados: riesgo relativo de padecer cáncer de pulmón por exposición a sílice en no silicóticos: 0,96; IC 95% (0,81-1,15). Riesgo relativo de padecer cáncer de pulmón por exposición a sílice en silicóticos: 2,37; IC 95% (1,98-2,84). Usted concluye que:

- 1) La exposición al sílice es un factor de riesgo (FR) de cáncer independientemente de si el sujeto es silicótico.
- 2) La exposición al sílice es protectora del cáncer en los no silicóticos.
- 3) En los silicóticos, la exposición al sílice es un FR de cáncer.
- 4) Los intervalos están mal calculados por ser asimétricos respecto al valor muestral.

RC: 3

La incidencia de hepatocarcinoma en un determinado país, en sujetos alcohólicos, es de 10 casos nuevos por cada 1.000 bebedores. Sin embargo, el riesgo de hepatocarcinoma es de 2 por cada 2.000 no bebedores. ¿Cuánto más riesgo de hepatocarcinoma tienen los alcohólicos con respecto a los no alcohólicos?

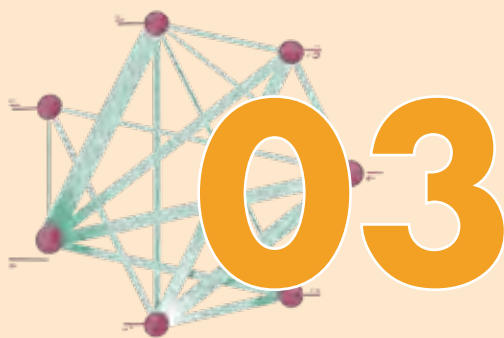
- 1) 1.
- 2) 10.
- 3) 100.
- 4) 1.000.

RC: 2

En un ensayo clínico que compara un anorexigénico frente a placebo en la reducción de IAM, se ha comprobado un riesgo de un 5% de infarto en los que tomaban el fármaco, mientras que era del 10% en los que recibieron placebo. Calcule cuántos sujetos deben tomar el fármaco para evitar un IAM:

- 1) 2.
- 2) 5.
- 3) 10.
- 4) 20.

RC: 4



Validez y fiabilidad de los estudios epidemiológicos

Orientación MIR

Suelen presentarse estudios en los que se ha cometido algún error sistemático y puede no ser sencillo identificar el posible sesgo que se ha cometido. El tipo de sesgo más preguntado es el de confusión. De este, hay que saber en qué consiste y cómo se puede evitar.

Preguntas MIR

- ▶ MIR 23-24, 45
- ▶ MIR 22-23, 44; MIR 22-23, 49
- ▶ MIR 21-22, 45
- ▶ MIR 16-17, 121
- ▶ MIR 13-14, 197; MIR 13-14, 198; MIR 13-14, 200

Conceptos clave

- ⊙ Los errores aleatorios son errores producidos por el azar. Son una falta de precisión, no afectan a la validez interna y se corrigen aumentando el tamaño de la muestra.
- ⊙ Los errores sistemáticos o sesgos no se deben al azar, afectan a la validez de los estudios y no se corrigen aumentando el tamaño de la muestra.
- ⊙ Los sesgos de selección se deben al hecho de elegir mal la muestra del estudio, bien por escoger controles hospitalarios (Berkson), por disponer de pocos casos enfermos al estudiar enfermedades de rápida mortalidad (Neyman), o por seleccionar voluntarios.
- ⊙ Los errores de clasificación se deben a un error al incluir a los pacientes en los diferentes grupos de estudio, ya sea porque los sanos (controles) no recuerdan el antecedente de la exposición (sesgo de memoria), o porque varían su comportamiento al sentirse observados.
- ⊙ En los ensayos clínicos se pueden controlar los sesgos de clasificación mediante las técnicas de ciego o enmascaramiento; y los de selección, por medio de la aleatorización.
- ⊙ El sesgo de confusión se debe a la existencia de un factor de confusión, que es esencialmente un factor que está desigualmente distribuido entre los dos grupos de pacientes del estudio, y que se relaciona etiológicamente con la enfermedad final. Los sesgos de confusión se controlan mediante asignación aleatoria, apareamiento, restricción, análisis multivariante y análisis estratificado.

Cuando se realiza un estudio de investigación clínica, casi nunca se trabaja con poblaciones completas. Lo que se hace es, partiendo de observaciones realizadas en un grupo reducido de personas (la llamada muestra), generalizar o extrapolar los resultados obtenidos a colectivos más amplios. El hecho de que no se trabaje con poblaciones completas, sino con muestras, puede introducir en las observaciones errores producidos por el azar, a los que se les llama errores aleatorios. Existen además otros tipos de errores no relacionados con el hecho de trabajar con muestras, y que son conocidos como errores sistemáticos o sesgos.

3.1. Tipos de error

Los tipos de error son los siguientes:

- **Error aleatorio.** Es el error que puede atribuirse a la variabilidad aleatoria que conlleva siempre un proceso de muestreo. El azar hace que la muestra con la que se va a trabajar no sea representativa. El error aleatorio no afecta a la validez interna de un estudio, pero reduce la posibilidad de elaborar conclusiones sobre la relación exposición-enfermedad, aunque no altera el sentido de la asociación. Los errores aleatorios, a diferencia de los errores sistemáticos, se pueden minimizar aumentando el tamaño de la muestra.
- **Errores sistemáticos o sesgos.** Son los errores producidos cuando hay un fallo en el diseño, o en la ejecución del estudio, que hace que los resultados de la muestra sean diferentes de la población de la que proceden. No se relacionan con el tamaño de la muestra y, cuando no se controlan, tienden a invalidar las condiciones de un estudio; es decir, la existencia de sesgos conduce a la elaboración de conclusiones incorrectas sobre la relación entre una exposición y una enfermedad

► MIR 22-23, 44; MIR 21-22, 45.

Recuerda

- La mayoría de los sesgos acontecen habitualmente en los estudios de casos-controles (retrospectivos).

Tipos de errores sistemáticos

Sesgo de selección

Se produce cuando se asignan sujetos al grupo control que difieren significativamente, en alguna característica clave, del grupo problema. Este tipo de sesgos se puede controlar mediante un proceso de aleatorización en la constitución de los distintos grupos de estudio.

Son ejemplos de este tipo de sesgo los siguientes:

- **Sesgo de autoselección o del voluntario.** La participación o autoderivación del individuo al estudio compromete su validez.
- **Sesgo diagnóstico o sesgo de Berkson.** Es el que se produce cuando, para saber qué ocurre en la población, se elige una muestra hospitalaria de esa población y el factor de riesgo que se está estudiando se asocia a una mayor probabilidad de hospitalización. También se incluye en este tipo de sesgo aquel que puede surgir al seleccionar como control a pacientes con alguna enfermedad que también se asocia al factor de exposición estudiado.

- **Sesgo del obrero sano.** La salida del trabajador enfermo del mercado laboral compromete la validez del estudio.
- **Falacia de Neyman.** Se produce en los estudios de casos-controles al seleccionar casos prevalentes (ya existentes) en vez de casos incidentes (casos nuevos). Esto conlleva que en los casos sea menos frecuente la exposición a aquellos FR que disminuyen la supervivencia; esto es, los casos representarían a individuos con rasgos de mayor resistencia a la enfermedad o formas menos graves de esta. Por ejemplo, se quiere estudiar si la actividad física tiene algún efecto sobre la frecuencia de la insuficiencia coronaria. Se compara a personas con actividad y a personas sedentarias. Un sesgo de selección sería cuando la inactividad de los sedentarios fuese a consecuencia de padecer la enfermedad cardíaca.

Sesgo de información

Tiene lugar cuando hay un error sistemático en la medición de alguna variable clave del estudio. La clasificación incorrecta puede o no afectar de igual manera a todos los grupos de estudio. Los sesgos de información debidos al investigador o a los participantes en el estudio se controlan mediante técnicas de enmascaramiento.

- **Clasificación incorrecta diferencial.** La probabilidad de clasificación errónea de la exposición (en un estudio de casos-controles) afecta de forma desigual a sanos y enfermos o la clasificación de enfermo o no enfermo (en un estudio de cohortes) se hace de manera distinta, según se esté o no expuesto al factor de estudio. Este tipo de sesgo produce una subestimación o sobreestimación del efecto del factor de exposición en la enfermedad. Dentro de este grupo cabe destacar:
 - **Sesgo de memoria.** Se produce en los estudios de casos-controles. El hecho de padecer la enfermedad hace que se esté más motivado para recordar posibles antecedentes de exposición ► MIR 13-14, 198.
 - **Sesgo de atención o efecto Hawthorne.** Los participantes en un estudio pueden modificar su comportamiento si saben que están siendo observados.
 - **Sesgo de regresión a la media.** Consiste en que una variable puede ser extrema en su primera medición por motivos fisiológicos, pero volver a la normalidad posteriormente. Por ejemplo, efecto de la bata blanca, se clasifica al sujeto expuesto, cuando realmente no lo es.
- **Clasificación incorrecta no diferencial.** La probabilidad de clasificación errónea ocurre en todos los grupos de estudio de forma similar. El error de clasificación no diferencial ocasiona una infraestimación (hace que la medida de asociación tienda a 1) del efecto del factor de exposición estudiado en la enfermedad ► MIR 22-23, 49; MIR 16-17, 121. Por ejemplo, se quiere estudiar el efecto del tabaco sobre la aparición de bronquitis crónica, comparándose a un grupo de fumadores y a otro de no fumadores. Existe la posibilidad de que los fumadores sean diagnosticados más fácilmente de bronquitis crónica que los no fumadores, simplemente por el hecho de que fumar se considera que está asociado a esta patología (Tabla 3.1).

Factor de confusión (confounding)

Un **factor de confusión (FC)** es una variable extraña al estudio que modifica los resultados que se obtienen ► MIR 13-14, 200. Todo FC debe cumplir tres condiciones:

- Ser un factor de riesgo para la enfermedad.
- Estar asociado con la exposición.
- No ser un paso intermedio entre la exposición y la enfermedad.



TIPO DE ERROR	CARACTERÍSTICAS	AFECTA A LA VALIDEZ INTERNA	SOLUCIÓN O CONTROL
Error aleatorio	Simétrico Incorregible Impredecible Estimable por estadística	No	↑ tamaño muestral
Sesgo o error sistemático	Asimétrico Corregible Predecible Prevención y control por epidemiología	Sí	Sesgo de selección: aleatorización Sesgo de clasificación: enmascaramiento Factor de confusión: • Aleatorización • Apareamiento • Restricción • Estratificación • Análisis multivariante

Tabla 3.1. Tipos de error.

Para prevenir los sesgos de confusión, existen distintas técnicas:

- **Fase de diseño.** Aleatorización (estudios experimentales), apareamiento y restricción.
- **Fase de análisis estadístico.** Análisis estratificado (se dividen los datos globales en dos grupos, según la presencia o no del factor de confusión) y análisis multivariante. El análisis multivariante será desarrollado en la parte de Estadística cuando se expliquen los modelos de regresión.

En la **Figura 3.1** se expone un ejemplo de análisis estratificado en el que se estudia la relación entre el tabaco y el cáncer de esófago, pero se sospecha que el alcohol puede ser un factor confusor. Se estratifica según los sujetos tengan o no el supuesto FC (alcohol en el ejemplo).

Tras la estratificación pueden obtenerse las tres situaciones que se muestran en el gráfico.

	Cáncer	No cáncer	
Fuma			
No fuma			
			100

Toma alcohol		OR cruda: 2,5		No toma alcohol	
	Cáncer	No cáncer		Cáncer	No cáncer
Fuma				Fuma	
No fuma				No fuma	
		30			70
OR toma alcohol				OR no toma alcohol	
2,5	Alcohol no FC		2,5		
1,8	Alcohol FC		1,8		
3,6	Alcohol F. modificador		1,3		

Figura 3.1. Interpretación del análisis estratificado.

3.2. Validez y fiabilidad

Se definen los siguientes conceptos:

- **Validez.** El estudio cuantifica lo que realmente se propone medir. Es el grado de ausencia de error sistemático. También recibe el nombre de exactitud.
- **Validez interna.** Consiste en el grado de validez del resultado para los pacientes del estudio. Se dice que un estudio tiene validez interna cuando los resultados del mismo son aplicables a los individuos del citado estudio ▶ **MIR 13-14, 197**.
- **Validez externa.** Un estudio tiene validez externa cuando los resultados del mismo son aplicables a otros individuos distintos de los del estudio.
- **Fiabilidad.** Es el grado de reproducibilidad de un estudio, esto es, el grado de similitud que presentarían los resultados si se repitiese el estudio en condiciones similares. En otras palabras, es el grado de ausencia de error aleatorio.

3.3. Validez de un test diagnóstico

Para conocer si una prueba diagnóstica es útil, se comparan sus resultados con los de otra prueba que actúa como patrón de referencia (*Gold standard*).

El grado en el que las medidas de la prueba evaluada se correlacionan con las medidas de referencia, se denomina **validez de criterio** (**Figura 3.2**).

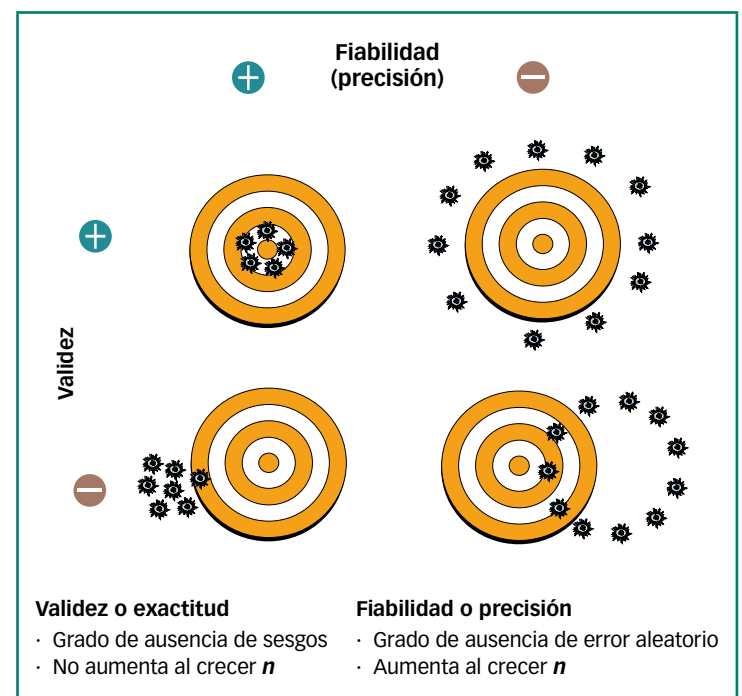


Figura 3.2. Validez y fiabilidad ▶ **MIR 23-24, 45**.

La prueba de referencia debe haber sido aceptada como tal por la comunidad científica; se ha de aplicar a toda la serie de casos estudiados y no puede incorporar información procedente de la prueba que evalúa.

El valor real de la prueba solo podrá ser establecido si el estudio se lleva a cabo en condiciones semejantes a la práctica clínica habitual, es decir, incorporando un amplio espectro de pacientes a los que en condiciones normales se les aplicaría dicho procedimiento diagnóstico.

Casos clínicos

En un estudio de casos-controles se encontró una asociación entre la ingesta de alcohol y el desarrollo de cáncer esofágico. Dado que el consumo de tabaco se encontró asociado tanto a la ingesta de alcohol como al desarrollo de cáncer de esófago, el consumo de tabaco en este estudio debe considerarse como:

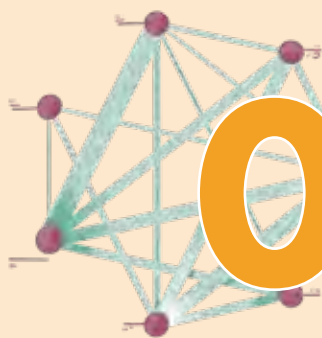
- 1) Un factor (sesgo) de confusión.
- 2) Un sesgo de información.
- 3) Un sesgo de memoria.
- 4) Un sesgo de selección.

RC: 1

En un estudio de cohortes se compara la eficacia de dos terapias antirretrovirales en la reducción de la carga viral de VIH. Para saber si la coinfección por virus C puede ser un factor que determine la respuesta antirretroviral, usted no deberá hacer:

- 1) Aleatorización.
- 2) Apareamiento.
- 3) Restricción.
- 4) Análisis multivariante.

RC: 1



Tipos de estudios epidemiológicos

Orientación MIR

Este es, sin duda, el tema más importante de la asignatura, y probablemente de todo el MIR. Hay dos tipos de preguntas características. Por una parte, las que se refieren a temas concretos o aspectos teóricos de los diferentes tipos de estudios. Por otra, un tipo de preguntas más "clínico": a partir de un estudio, se pide que se indique de qué tipo de estudio se trata.

Preguntas MIR

- ▷ MIR 23-24, 47
- ▷ MIR 22-23, 43
- ▷ MIR 21-22, 41; MIR 21-22, 48
- ▷ MIR 20-21, 45
- ▷ MIR 19-20, 30
- ▷ MIR 18-19, 120; MIR 18-19, 121
- ▷ MIR 17-18, 211; MIR 17-18, 212; MIR 17-18, 213; MIR 17-18, 224
- ▷ MIR 16-17, 120; MIR 16-17, 131; MIR 16-17, 132; MIR 16-17, 133
- ▷ MIR 15-16, 192; MIR 15-16, 193
- ▷ MIR 14-15, 182; MIR 14-15, 183; MIR 14-15, 197
- ▷ MIR 13-14, 193; MIR 13-14, 199

Conceptos clave

- ⊙ Los estudios experimentales son aquellos en los que el investigador forma parte "activa" del estudio al asignar el factor de estudio. La aleatorización de la muestra es otra condición indispensable.
- ⊙ Los estudios observacionales son aquellos en los que el investigador se limita a "observar" a los sujetos enfermos-sanos (estudio de casos-contróles), a sujetos expuestos-no expuestos (cohórtes), o a observar cuántos sujetos están enfermos en una comunidad (transversal).
- ⊙ Los estudios de cohórtes son prospectivos porque se empieza a partir del factor de riesgo (FR hacia enfermedad). Es decir, que a tiempo 0, nadie presenta el evento final de estudio. Son los mejores estudios de entre los observacionales para verificar la causalidad. La medida de asociación es el RR.
- ⊙ Los estudios de casos-contróles son retrospectivos porque la selección de pacientes se hace en función de si están enfermos o no (enfermedad hacia FR). En este estudio a tiempo 0, ya hay sujetos portadores del evento final de estudio. Son los mejores para enfermedades raras y su principal limitación es la relativa validez por la facilidad para cometer sesgos. La medida de asociación es la OR.
- ⊙ Los estudios transversales o de prevalencia se caracterizan por la ausencia de seguimiento. En estos estudios solamente "se ve" a los sujetos una única vez. La medida de frecuencia de enfermedad es la prevalencia.

4.1. Tipos de estudio

En la **Tabla 4.1.** se resumen los tipos de estudio:

FINALIDAD DEL ESTUDIO	Descriptivo	Describir una característica o enfermedad en una población
	Analítico	Evaluar presunta relación causa-efecto
CONTROL DE LA ASIGNACIÓN DEL FE	Experimental	Asignación controlada por el investigador
	Observacional	No hay asignación por parte del investigador
SECUENCIA TEMPORAL	Transversal	Evaluación en un tiempo concreto de la relación FE-E
	Longitudinal	Existe un tiempo entre las distintas variables. Puede establecerse una secuencia temporal entre estas. Pueden ser prospectivos o retrospectivos
RELACIÓN TEMPORAL FE-E	Prospectivo	Desde el FE hacia la E
	Retrospectivo	Desde la E hacia el FE

FE: factor de estudio; E: enfermedad.

Tabla 4.1. Tipos de estudios epidemiológicos.

4.2. Estudios descriptivos

Los objetivos de los estudios descriptivos son los siguientes:

- Describir las características y la frecuencia de un problema de salud, en función de las características de la persona (edad, sexo, estado civil...), del lugar (área geográfica...) y del tiempo de aparición del problema y su tendencia ► **MIR 14-15, 197**.
- Servir de base para estudios analíticos.

■ Series de casos clínicos

Describen las características de un grupo de enfermos. Son estudios longitudinales ya que contienen información adquirida a lo largo del tiempo. Su principal ventaja es que permiten generar nuevas hipótesis, mientras que el mayor inconveniente es que no presentan grupo control, por lo que cualquier FR puede ser un hallazgo casual ► **MIR 17-18, 213**.

■ Estudios ecológicos

Los conglomerados pueden estar constituidos por grupos poblacionales, comunidades, regiones, o países. La característica principal de este tipo de estudios es que se cuenta con información sobre la exposición o el evento para el conglomerado en su totalidad, desconociéndose la información a nivel individual para cada uno de los miembros del grupo. Son estudios en los que la unidad de análisis son grupos de individuos, no individuos (p. ej., clases de una escuela, ciudades, regiones).

Son útiles cuando no se puede hacer mediciones de exposición individuales (contaminación del aire, ruidos...). En este tipo de estudios es común asignar la misma exposición (exposición promedio) a todo el conglomerado, no considerando la variación individual de cada miembro del mismo. Igual sucede con la medición del evento; dado que solo se cuenta con el número de eventos registrados para el conglomerado, no es posible

discernir los eventos que se presentaron en los sujetos expuestos de los que ocurrieron en los no expuestos, por lo que es frecuente atribuir la totalidad de eventos a la exposición promedio que se registró en el interior del conglomerado.

Los estudios ecológicos permiten estudiar grandes grupos poblacionales en poco tiempo y con un coste relativamente muy bajo ya que, en general, utilizan estadísticas existentes recolectadas con otros fines. Sin embargo, dado que son uno de los peores estudios en la escala de causalidad, deben ser considerados únicamente para sugerir hipótesis, que tendrán necesariamente que ser verificadas mediante otros estudios más rigurosos de base individual, realizados posteriormente ► **MIR 21-22, 41; MIR 15-16, 192**.

Ventajas

Permiten describir diferencias en poblaciones que habrán de ser estudiadas con más detalle posteriormente.

Limitaciones

Los datos son promedios de poblaciones. Se usan medidas aproximadas de exposición (impuestos por alcohol, ventas de cajetillas de cigarrillos...) y de enfermedad (mortalidad en vez de incidencia...), lo que limita el valor de los hallazgos. Los principales problemas de este tipo de estudios son que se ignora la variabilidad individual de los integrantes de los conglomerados y que no es posible corregir por diferencias en otras variables que pudieran estar también asociadas con la exposición y el evento en estudio.

■ Estudios transversales o de prevalencia

Son estudios descriptivos y transversales, ya que analizan la relación entre una enfermedad y algunas variables en un momento concreto del tiempo. Buscan hallar una posible relación entre un FR y una enfermedad, que luego habrá de ser verificada por estudios analíticos ► **MIR 15-16, 193**.

Características

Es de "corte" o transversal, ya que enfermedad y características se miden simultáneamente.

Ventajas

No tienen problemas éticos, son de duración mínima, su coste es bajo y son de fácil reproducibilidad. Son útiles para el estudio de enfermedades crónicas en la población.

Inconvenientes

No es útil para estudiar enfermedades raras, ya que no permite ver el mecanismo de producción de la enfermedad y no sirve para comprobar una hipótesis previa de causalidad (es imposible conocer la secuencia temporal porque la información sobre el factor de riesgo y la enfermedad se recogen a la vez).



■ Análisis de las medidas de enfermedad

Con el estudio de prevalencia, la medida que se obtiene es la razón de prevalencia de individuos expuestos.

4.3. Estudios analíticos

Los estudios analíticos intentan establecer una relación de causalidad entre el factor de riesgo y la enfermedad. Se pueden clasificar en:

- Experimentales.
- Observacionales.

En los primeros es el investigador el que asigna el factor de estudio (qué fármaco, vacuna, campaña de educación..., cuánto tiempo, cuándo, cuánta dosis recibirán los individuos...), mientras que en los segundos, el investigador se limita a observar qué es lo que sucede en un grupo de individuos, sin manipular el estudio.

■ Estudios analíticos experimentales

Se dice que un estudio es experimental cuando cumple estas condiciones:

- Asignación por parte del investigador del factor de estudio.
- Aleatorización de la muestra de modo que los participantes son adscritos al azar a uno u otro grupo de estudio.

Limitaciones. Los problemas éticos son el principal inconveniente de este tipo de estudios. No es admisible exponer a unos sujetos a un factor de riesgo que presumiblemente es el causante de una enfermedad.

Ventajas:

- Son los estudios que mejor valoran la utilidad de una intervención y aportan mayor evidencia causal.
- Permiten un gran control sobre cualquier efecto que no sea el estudiado.
- Posibilitan el empleo de técnicas de enmascaramiento.

Hay dos tipos de estudios analíticos experimentales son los ensayos clínicos aleatorios y los ensayos de campo.

Ensayo clínico aleatorio

Es, con mucho, el estudio experimental más frecuente. La asignación aleatorizada del factor de estudio (un fármaco o una intervención sanitaria) se hace sobre los individuos. Es el mejor para demostrar causalidad y la eficacia de una actuación. Por ejemplo, la asignación aleatorizada de un antirretroviral o un placebo a un grupo de pacientes con VIH ► MIR 13-14, 193 (para ampliar el estudio, véase el tema 6).

Recuerda

- No siempre que se evalúan dos tratamientos se está ante un ensayo clínico. Si los pacientes no fueron asignados por el investigador y este no introdujo el factor de estudio, se estaría ante un estudio de cohortes.

Ensayo de campo

Es un estudio que valora la eficacia de una medida preventiva. En general, estos estudios son más caros que los ensayos clínicos y requieren mayor número de individuos.

Las principales diferencias respecto a los ensayos clínicos son:

- Se hacen sobre individuos sanos.
- Valoran la eficacia de las medidas preventivas.

Por ejemplo, la evaluación de la eficacia de una campaña de vacunación en 10.000 niños, de forma que 5.000 reciben aleatoriamente una vacuna y los otros 5.000, un placebo.

■ Estudios analíticos cuasiexperimentales

Se diferencian de los estudios experimentales puros en que no hay asignación al azar (aleatorización).

Los tipos de estudios analíticos cuasiexperimentales son: los ensayos comunitarios de intervención, los ensayos antes-después y los estudios controlados no aleatorios.

Ensayo comunitario de intervención

Es una variedad de los ensayos de campo:

- Se trabaja con individuos sanos.
- Valora la eficacia de medidas preventivas.
- No se aplica aleatorización individual.

Por ejemplo, una campaña sanitaria de agua fluorada en una comunidad y de agua clorada en la otra, comparando la frecuencia de caries en los dos grupos.

Ensayos antes-después

En este tipo de estudios, el fármaco (o medida en general) se administra a los individuos y se compara el resultado con la situación basal. Los estudios antes-después tienen la ventaja de que son más fáciles de hacer, pero poseen el inconveniente grave de que, al no disponer de grupo de control, los resultados son difíciles de interpretar.

Por ejemplo, a un grupo de pacientes cardiopatas se les administra un IECA, comparando la presión arterial (PA) media en el conjunto antes y después de administrar el tratamiento.

Estudios controlados no aleatorios

Se realizan cuando la asignación aleatoria no ofrece ventajas o no se puede hacer. Por ejemplo, se quiere hacer un ensayo en pacientes cardiopatas. A un grupo se le administra IECA (se trata de pacientes no dislipidémicos) y al otro (que además son dislipidémicos), IECA más estatinas.

En este tipo de estudios no hay asignación aleatorizada, sino que la inclusión en uno u otro grupo de estudio se ha hecho tomando como base los factores de riesgo que presentan los pacientes.

Estudios analíticos observacionales

Los tipos de estudios analíticos observacionales son los que se enumeran a continuación.

Estudio de cohortes

Partiendo de un grupo de individuos expuestos al FR (cohorte expuesta) (Figura 4.1), y de otro conjunto comparable en todo pero cuyos individuos no están expuestos al FR (cohorte no expuesta), se estudia la incidencia de la enfermedad en ambas cohortes.

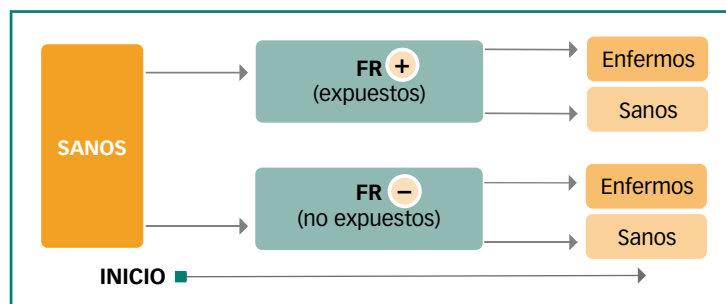


Figura 4.1. Estudio de cohortes.

Por ejemplo, se sigue a un grupo de fumadores y a otro de no fumadores a lo largo del tiempo, observando el número de casos nuevos de cáncer de pulmón que aparecen en cada grupo de pacientes ▶ MIR 22-23, 43; MIR 21-22, 48; MIR 20-21, 45; MIR 19-20, 30; MIR 18-19, 121; MIR 17-18, 211; MIR 17-18, 212; MIR 14-15, 182.

Características

A continuación, se enumeran las características de los estudios de cohortes:

- Son estudios longitudinales, de seguimiento.
- Es prospectivo (excepto en los estudios de cohortes históricas).
- Va de la causa al efecto (enfermedad).

Ventajas

Las ventajas se pueden resumir en que:

- Es el mejor estudio para comprobar hipótesis previas de causalidad cuando, por razones éticas, no es posible realizar un estudio experimental.
- Es el mejor para el estudio de la "multiefectividad del factor de riesgo" (todos los efectos del factor de riesgo).
- La posibilidad de sesgos es baja.
- Sirve para el estudio de exposiciones raras.

Inconvenientes

Los estudios de cohortes cuenta también con inconvenientes:

- No es bueno para el estudio de enfermedades raras.
- No es adecuado para el estudio de enfermedades de largo período de incubación.
- El coste es alto.
- No sirve para el estudio de la "multicausalidad de la enfermedad".
- Es difícilmente reproducible.

Análisis de las medidas de la enfermedad

Los estudios de cohortes son los que permiten saber cuál es la incidencia de la enfermedad. Las medidas que se obtienen son:

- **Riesgo relativo.** Es la medida de la fuerza de la asociación.
- **Diferencia de incidencias o riesgo atribuible.** Informa sobre el exceso de riesgo de enfermar.
- **Fracción atribuible.** Estima la proporción de la enfermedad entre los expuestos que es debida al factor de riesgo.

Recordar

- Es necesario darse cuenta de que como en un estudio de cohortes todos los sujetos están sanos en el momento inicial, los enfermos que vayan apareciendo serán nuevos, es decir, incidentes. Por ello, la medida de frecuencia es la incidencia y la de asociación, el riesgo relativo.

Estudio de cohortes históricas (retrospectivas)

En esta variante de el estudio de cohortes, el investigador identifica, mediante registros, una cohorte expuesta en el pasado a un factor de riesgo, y otra cohorte no expuesta. Mediante dichos registros (p. ej., historias clínicas) sigue la evolución de ambas cohortes, comparando los resultados.

Por ejemplo, se selecciona una muestra de 100 historias clínicas de un hospital que se divide en dos grupos en función del antecedente de tabaquismo o no. Se observa en cada grupo la frecuencia de cáncer de laringe

▶ MIR 23-24, 47.

Estudio de casos-contróles

Partiendo de un grupo de individuos enfermos (casos) y de otro comparable a ellos en todo, pero que no tienen la enfermedad (contróles) (Figura 4.2), se estudia la exposición, en ambos casos, a distintos factores de riesgo. Por ejemplo, se comparan los factores de riesgo a los que ha estado expuesto un grupo de pacientes con cáncer de páncreas frente a los factores a los que ha estado expuesto un grupo de individuos sanos ▶ MIR 17-18, 224; MIR 16-17, 120; MIR 16-17, 131; MIR 16-17, 133; MIR 14-15, 183; MIR 13-14, 199.

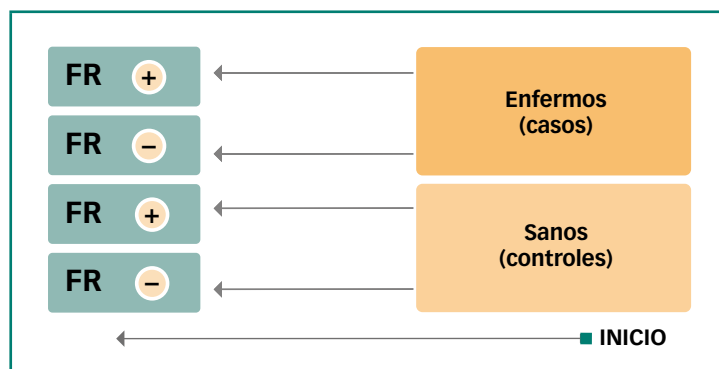


Figura 4.2. Estudio de casos-contróles.



● Características

A continuación, se enumeran las características de los estudios de casos-contróles:

- Es un estudio longitudinal.
- Es retrospectivo.
- Va del efecto (enfermedad) a la causa.

● Ventajas

Las ventajas de este tipo de estudios se resumen en que:

- Es de corta duración.
- Es ideal para el estudio de enfermedades raras.
- Es el mejor para el estudio de enfermedades de largo período de inducción.
- El coste es bajo.
- Es el más adecuado para el estudio de la multicausalidad de la enfermedad (los posibles factores de riesgo de una determinada enfermedad).
- Es el mejor para formular nuevas hipótesis etiológicas.

● Inconvenientes

Existen tres inconvenientes:

- No es adecuado para comprobar hipótesis previas de causalidad.
- No permite el estudio de la “multiefectividad del factor de riesgo”.
- La posibilidad de sesgos es muy alta, su control difícil.

● Análisis de las medidas de la enfermedad

En los estudios de casos-contróles no es posible obtener información sobre la incidencia de la enfermedad, ya que se parte de una población seleccionada. Tampoco se tiene información acerca de la prevalencia, pues el

número de pacientes solo depende de los que se elijan. Debido a ello, la fuerza de la asociación no se puede calcular directamente, como en el estudio de cohortes, sino de forma indirecta mediante la *odds ratio*.

Casos-contróles anidados

En esta variante el tipo de estudio de casos-contróles, la serie de controles está muestreada aleatoriamente de la cohorte que da origen a los casos; de ahí su nombre de casos-contróles anidados **MIR 18-19, 120; MIR 16-17, 132**. De esta forma, al originarse tanto casos como controles de una cohorte común, se minimiza la probabilidad de sesgos.

En la **Tabla 4.2** se exponen las diferencias fundamentales entre los dos tipos de estudios analíticos observacionales, de cohortes y de casos-contróles.

CARACTERÍSTICAS	ESTUDIO DE COHORTES	ESTUDIO DE CASOS-CONTRÓLES
Población en riesgo	Definida	Indefinida
Cálculo de la incidencia	Sí	No
N.º de efectos medibles	Varios	Uno
N.º de exposiciones posibles a estudiar	Una	Varias
Enfermedades raras	Poco útil	Muy útil
Exposiciones raras	Muy útil	Poco útil
Sesgos posibles	Pocos	Muchos
N.º de sujetos en estudio	Muchos	Pocos
Coste	Elevado	Bajo
Duración	Larga	Corta
Evidencia causal	Regular	Mala

Tabla 4.2. Diferencias entre los estudios de cohortes y los de casos-contróles.

Casos clínicos

Para tratar de establecer una relación causal entre el consumo de talidomida durante el embarazo y el riesgo de focomelia, se seleccionaron madres de recién nacidos con focomelia y se compararon con madres de recién nacidos sanos, en cuanto a los antecedentes de toma de talidomida. ¿Cuál es el tipo de diseño de estudio empleado?

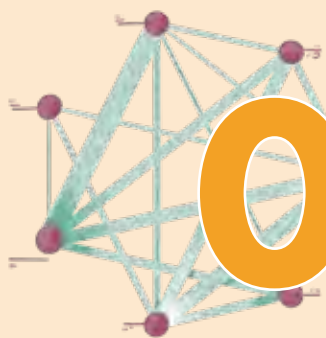
- 1) Casos-contróles.
- 2) Estudio de cohortes.
- 3) Ensayo clínico aleatorizado.
- 4) Estudio ecológico.

RC: 1

Se realiza un muestreo aleatorio de historias clínicas para estudiar la relación entre tabaco y cáncer de pulmón. Dichas historias clasifican inicialmente según el antecedente personal de tabaquismo. ¿De qué estudio se trata?

- 1) Casos-contróles.
- 2) Estudio de cohortes.
- 3) Ensayo clínico aleatorizado.
- 4) Estudio ecológico.

RC: 2



El ensayo clínico

Orientación MIR

Como ya se ha comentado anteriormente, uno de los temas más importantes del MIR es el ensayo clínico. En el examen preguntan todo tipo de aspectos, sin embargo, lo más habitual es que o bien pregunten conceptos teóricos, o bien pongan un ensayo clínico y pregunten de qué tipo de ensayo se trata.

Preguntas MIR

- ▶ MIR 23-24, 49; MIR 23-24, 50; MIR 23-24, 51
- ▶ MIR 22-23, 52
- ▶ MIR 20-21, 48
- ▶ MIR 18-19, 116; MIR 18-19, 132
- ▶ MIR 17-18, 25; MIR 17-18, 208-BL; MIR 17-18, 209; MIR 17-18, 210; MIR 17-18, 225; MIR 17-18, 228
- ▶ MIR 16-17, 36; MIR 16-17, 116; MIR 16-17, 117; MIR 16-17, 119
- ▶ MIR 15-16, 29; MIR 15-16, 38; MIR 15-16, 190
- ▶ MIR 14-15, 178; MIR 14-15, 179; MIR 14-15, 180; MIR 14-15, 190; MIR 14-15, 193; MIR 14-15, 196
- ▶ MIR 13-14, 202; MIR 13-14, 203

Conceptos clave

- ⊙ El objetivo de la fase I del EC es determinar las propiedades farmacodinámicas del fármaco, así como la dosis tóxica del mismo.
- ⊙ En una fase II se establece la dosis eficaz del fármaco.
- ⊙ La fase III se caracteriza porque se compara la eficacia de otro fármaco con la del que se está investigando. Tras esta fase, el fármaco se introduce en el mercado.
- ⊙ La fase IV es la de farmacovigilancia. En este caso, se registra la aparición de efectos secundarios infrecuentes. La notificación en la práctica clínica de posibles eventos adversos se realiza mediante la tarjeta amarilla.
- ⊙ La secuencia correcta en un EC es:
 - Los pacientes cumplen los criterios de inclusión y exclusión.
 - Firman el consentimiento informado (sabiendo que posiblemente reciban placebo, si uno de los tratamientos lo es).
 - Son asignados a uno u otro grupo de intervención (preferiblemente mediante aleatorización).
- ⊙ Un aspecto importante es manejar las pérdidas que se producen en un EC. Lo mejor es tenerlas previstas a priori. Si aun así se produjeran, sería conveniente realizar un nuevo estudio con el tamaño apropiado. Sin embargo, no todas las pérdidas afectan de igual manera al EC. Si las pérdidas se producen antes de la asignación del factor de estudio, no necesariamente se ve afectada la validez del mismo, al contrario que si se producen después de tal asignación.
- ⊙ Los EC se dividen en:
 - Pragmáticos (muestra heterogénea) y explicativos (muestra homogénea).
 - Cruzados (cada grupo de sujetos toma los dos fármacos con un tiempo de lavado) y paralelos (cada grupo de tratamiento toma un único fármaco).
 - Por intención de tratar (todos los sujetos incluidos en el estudio son analizados) y por protocolo (solo se analizan los que finalizan el estudio).
 - De tamaño predeterminado (se conoce a priori el tamaño de la muestra) o secuenciales (los pacientes van incorporándose al estudio progresivamente).
 - De superioridad (interesa demostrar que un fármaco es mejor que otro) y de no-inferioridad (se supone una eficacia similar, la ventaja del nuevo fármaco es ser más barato, tener una administración más fácil...).
 - Enmascarados (se han realizado técnicas de ciego para evitar sesgos) o abiertos (no enmascarados).

5.1. Introducción

Un **ensayo clínico (EC)** es un experimento cuidadosa y éticamente diseñado con el fin de dar respuesta a preguntas que tienen que ver con la utilidad de procedimientos diagnósticos, terapéuticos y profilácticos en el ser humano. La autorización de la Agencia Española de Medicamentos es necesaria para el desarrollo de cualquier EC con medicamentos en un centro sanitario, junto con un informe favorable del Comité Ético de Investigación Clínica del centro sanitario ► **MIR 17-18, 225**.

Habitualmente los EC se realizan en adultos inicialmente, dejando para posterior ocasión el estudio en los extremos de la vida ► **MIR 17-18, 208-BL**.

Los EC son estudios prospectivos y experimentales en los que, una vez seleccionada la muestra, se divide aleatoriamente en dos grupos de pronóstico comparable que idealmente solo se diferencian en la intervención terapéutica que van a recibir ► **MIR 22-23, 52**.

Teniendo en cuenta los objetivos perseguidos en el desarrollo de un medicamento, se distinguen cuatro fases (**Tabla 5.1**):

- **Ensayo en fase I.** Es la primera vez que un fármaco se administra a humanos. Generalmente, se realiza con voluntarios sanos ($n = 20 - 80$) y sin grupo control. Suelen ser EC abiertos, no aleatorizados y, muchas veces, no comparativos. El principal objetivo es evaluar la toxicidad y conocer la dosis única aceptable no tóxica. Se limitan al terreno experimental y/o a sus efectos farmacocinéticos y farmacodinámicos ► **MIR 14-15, 179**.
- **Ensayo en fase II.** El principal objetivo es aportar información sobre la relación dosis-respuesta, proporcionando una información preliminar acerca de la eficacia. Se lleva a cabo en pacientes ($n = 100 - 200$). No necesariamente tienen que ser comparativos. Generalmente, este tipo de EC es controlado y con asignación aleatoria del tratamiento. Algunas clasificaciones distinguen a su vez una fase IIa (estudios piloto realizados en pocos pacientes con criterios de inclusión/exclusión más estrictos) y una fase IIb (que evalúa la eficacia y la seguridad en un mayor número de pacientes y representa una demostración más rigurosa de la eficacia del nuevo compuesto) ► **MIR 23-24, 50; MIR 20-21, 48; MIR 16-17, 116; MIR 13-14, 203**.
- **Ensayo en fase III.** Es el prototipo del EC. Suele ser comparativo con la terapéutica de referencia o con un placebo. Es la investigación clínica más extensa y rigurosa sobre un tratamiento médico. Sirve para establecer la eficacia de un nuevo fármaco y la existencia de efectos adversos frecuentes. Un paciente debe conocer que probablemente ha recibido placebo, pero no si lo ha recibido con certeza o no. Lo ideal es comparar frente al mejor tratamiento, a la mejor alternativa existente. Se realiza en

un grupo mayor de pacientes (varias decenas de miles), seleccionados con unos criterios de inclusión/exclusión más laxos, que normalmente se encuentran polimedicados y que permitan hacer una extrapolación de los resultados al resto de la población (validez externa). Es posible, a su vez, distinguir las fases IIIa y IIIb. La primera incluye los EC que se llevan a cabo una vez determinada la eficacia terapéutica, pero antes de ser enviada a las autoridades reguladoras la solicitud de aplicación del nuevo compuesto, y constituyen la evidencia de efectividad terapéutica que se exige para la tramitación del expediente del nuevo compuesto. Con frecuencia son considerados como *pivotal trials*. La segunda fase (IIIb), comprende estudios que se realizan una vez aceptada la tramitación y antes de la aprobación y comercialización del nuevo compuesto.

- **Ensayo en fase IV.** Consiste en el seguimiento, posterior a la comercialización del fármaco, de un número de pacientes con el fin de detectar efectos adversos poco frecuentes o de aparición tardía. La fase IV también sirve para evaluar interacciones entre medicamentos en situaciones parecidas a las de la rutina clínica, lo que permite analizar indicaciones y efectos secundarios más o menos similares a la práctica habitual, y sugerir nuevas indicaciones de fármacos ya aceptados para otro fin.

Así, la farmacovigilancia es la recogida de datos sobre seguridad de fármacos una vez que se autoriza su comercialización ► **MIR 23-24, 49**. Su objetivo es detectar reacciones adversas poco frecuentes que no han sido identificadas durante el desarrollo del ensayo, dado que este se ha realizado en una población limitada (generalmente $n < 2.000$).

Pero los EC fase IV no son la única forma de evaluación poscomercialización. Se entiende por EPA (estudio posautorización) cualquier estudio realizado con medicamentos según su ficha técnica, en condiciones normales de uso. Como se ha comentado en ocasiones, son EC, pero lo habitual es que se trate de estudios observacionales en los que el investigador se limita a observar la realidad del fármaco en condiciones habituales de uso ► **MIR 16-17, 119**.

Los objetivos fundamentales de los EPA persiguen lograr, en condiciones reales:

- Determinar la efectividad del fármaco.
- Conocer la efectividad percibida por el paciente.
- Identificar factores de riesgo.
- Establecer nuevas indicaciones.

El Sistema Español de Farmacovigilancia recibe información de diferentes fuentes, entre ellas, y principalmente, de la notificación espontánea de reacciones adversas. Esta supone la recogida y la notificación, mediante las tarjetas amarillas, de las reacciones adversas que aparecen durante la práctica clínica.

	CARACTERÍSTICA PRINCIPAL	OBJETIVO	MUESTRA
Fase I	Primeros pasos en la investigación de un fármaco en seres humanos	Estimación inicial de tolerancia, seguridad, farmacocinética y farmacodinámica No estimación de eficacia	Habitualmente solo pocos voluntarios sanos ($n = 20 - 80$) No suelen tener grupo control ni enmascaramiento
Fase II*	Ensayo terapéutico, pero solo exploratorio	Información preliminar sobre eficacia, patrón dosis-respuesta, tolerancia y ampliación de los datos de seguridad obtenidos en fase I	Ensayos pequeños ($n = 100 - 200$) en pacientes con la enfermedad de interés Preferentemente con grupo control
Fase III	El más frecuente en publicaciones médicas Dirigido a lograr la aprobación o comercialización del fármaco	Demostrar un efecto terapéutico (eficacia) Valorar la seguridad del nuevo fármaco frente a alternativas disponibles	Muestra más amplia de pacientes con una única enfermedad concreta Controlados y aleatorizados
Fase IV	Vigilancia poscomercialización y a largo plazo	Evaluar seguridad y efectividad, nuevas indicaciones, efectos secundarios, morbilidad y mortalidad a largo plazo	Preferentemente controlados y aleatorizados

*Algunos autores dividen la fase II en IIa y IIb. En general, durante la fase IIa se realizan estudios piloto con pocos participantes y criterios de inclusión/exclusión muy estrictos. Se evalúa el efecto farmacológico y pueden participar tanto individuos enfermos como sanos. En la fase IIb se incluye un mayor número de participantes; el objetivo es medir el patrón dosis-respuesta y lo fundamental es valorar la eficacia, por lo que solo se incluyen sujetos enfermos.

Tabla 5.1. Tipos de ensayos clínicos según la fase de investigación.



En general, se acepta cualquier reacción adversa; sin embargo, las de mayor interés son las reacciones adversas graves, inesperadas o raras y aquellas de fármacos de comercialización reciente (menos de 3 años) ▶ **MIR 23-24, 51**. El programa de notificación espontánea está dirigido especialmente a médicos, aunque incluye a otros profesionales sanitarios.

5.2. Etapas en la realización de un ensayo clínico

En los siguientes apartados se desarrollan los pasos a seguir para la realización de un EC en fase III con dos grupos de intervención.

■ Selección de la cohorte de estudio

Consiste en la elección de un grupo de sujetos adecuados para la evaluación del medicamento, de los cuales se extrapolarán los resultados del EC. Mediante la formulación de los criterios de inclusión, se establecen las características clínicas y sociodemográficas de los pacientes en los que se empleará el fármaco. Los criterios de exclusión se usan para rechazar a aquellos pacientes con criterios de inclusión pero que presentan alguna contraindicación, condiciones que pueden afectar a la variable resultado o a alguna característica que haga complicado su estudio.

En función de los criterios de inclusión, los EC se pueden dividir en dos grandes grupos:

- **EC pragmáticos.** Se acercan lo más posible a la población general. Los criterios de inclusión son muy laxos (prácticamente los únicos son los diagnósticos) ▶ **MIR 14-15, 180**.
- **EC explicativos.** Los criterios de inclusión son muy estrictos, por lo que la muestra no es exactamente igual a la población general.

Si los criterios de inclusión son muy estrictos, el EC presenta una muestra más homogénea y se necesita menor tamaño muestral para detectar las diferencias. Además, hay mayor validez interna. Sin embargo, existe una clara limitación a la hora de generalizar los resultados y de reclutar a los pacientes. Se utilizan, sobre todo, en las fases II y III. En cambio, cuando no son estrictos, el reclutamiento se verá facilitado, se podrán generalizar los resultados con mayor facilidad y tendrá mayor validez externa. No obstante, la muestra será heterogénea, por lo que se necesita un mayor tamaño muestral y los resultados son confusos si solo es eficaz en subgrupos de pacientes. Se emplean en las fases III y IV.

En esta etapa también hay que determinar el tamaño muestral. Este ha de ser suficiente para obtener un IC del 95% de la eficacia cuyos límites sean clínicamente relevantes.

■ Medición de variables basales

Consiste en determinar una serie de variables en los sujetos que cumplen los criterios de inclusión y que han aceptado participar en el estudio (consentimiento informado), con los siguientes propósitos:

- Definir las características de la población a la que podrá extrapolarse el resultado del ensayo clínico.

- Aportar una serie de datos que permitan posteriormente verificar que la aleatorización ha sido eficaz y hacer el análisis estratificado (p. ej., por edad), en caso de diferencias entre los dos grupos.
- Demostrar que el evento resultado, cuya frecuencia se pretende disminuir con el fármaco estudiado, no está presente al comienzo del estudio.
- Registrar posibles predictores del evento interés, de modo que se pueda evaluar su interacción con el fármaco estudiado, si la aleatorización no los ha distribuido homogéneamente entre los distintos grupos de estudio.

■ Aleatorización

La aleatorización consiste en asignar por azar, sin que influya ningún factor, los pacientes de la muestra a los dos grupos de intervención, de modo que, si el tamaño muestral es suficientemente grande, se consiga una distribución homogénea de las variables predictoras en ambos grupos ▶ **MIR 17-18, 228**. Se puede realizar de tres formas diferentes:

- **Aleatorización simple.** Cada paciente tiene la misma probabilidad de ser asignado a cualquiera de los grupos de tratamiento. Con este método existe riesgo de desigualdad numérica entre grupos.
- **Aleatorización por bloques.** Se establecen bloques de aleatorización, de modo que en cada uno de ellos la mitad de los pacientes reciba el tratamiento experimental y la otra mitad, el control. Con este tipo de aleatorización se evita la desigualdad numérica entre los distintos grupos, por lo que es especialmente útil cuando el tamaño muestral no es muy grande.
- **Aleatorización estratificada.** Los pacientes son divididos en grupos (estratos) homogéneos respecto a alguna variable de interés pronóstico y posteriormente son asignados aleatoriamente a uno de los dos grupos de intervención (**Figura 5.1**) ▶ **MIR 17-18, 210**.

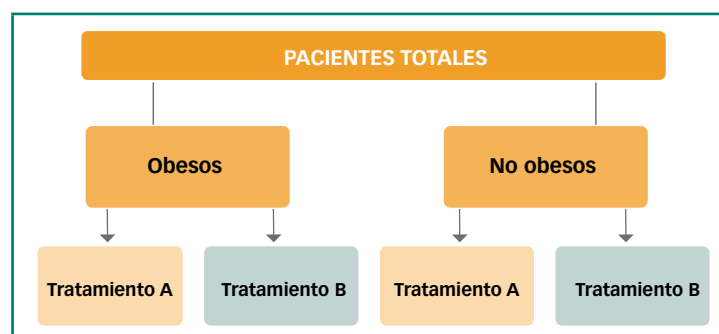


Figura 5.1. Aleatorización estratificada.

En general, la aleatorización debe conseguir asignar los tratamientos de forma desconocida e impredecible, tiene que ser un proceso reproducible y estar documentado; ha de basarse en propiedades matemáticas conocidas, no debe ser prevista la secuencia de los tratamientos y, finalmente, ha de ser posible detectar los fallos de la asignación. Para evitar sesgos en la aleatorización, es importante que quien decide la inclusión de los pacientes en el ensayo y la persona que lleva a cabo la aleatorización desconozcan la secuencia de aleatorización hasta que se aplique a cada uno de los pacientes reclutados. En caso de no hacerse ocultación de la secuencia de aleatorización, existe la posibilidad de seleccionar incorrectamente a los sujetos del ensayo y perder la comparabilidad de los dos grupos de tratamiento. La ocultación de la secuencia de aleatorización se puede conseguir mediante un sistema de aleatorización centralizado en un lugar distinto a aquel en el que se lleva a cabo el ensayo, o mediante sobres numerados secuencialmente, opacos y lacrados, que contengan el tratamiento y que solo se abran tras la inclusión del paciente en el EC.

■ Aplicación de la intervención

Es importante evitar que tanto investigadores como otros proveedores de cuidados traten de forma diferente a los dos grupos de intervención. Para evitar este hecho, llamado sesgo de cointervención, y que la medición de la variable-resultado pueda estar sesgada por el mismo motivo, se utiliza el enmascaramiento. Cuanto mejor se haga el enmascaramiento, más se atenuará el efecto placebo y menores serán las diferencias entre los grupos ► MIR 18-19, 116.

Los tipos de enmascaramiento son:

- **Simple ciego.** El paciente desconoce el grupo de tratamiento al que está asignado.
- **Doble ciego.** Paciente, personal sanitario-evaluador de la variable resultado lo desconocen ► MIR 12-13, 187; MIR 12-13, 189.
- **Triple ciego.** Además de los anteriores, el analista de los datos no sabe el tipo de tratamiento de cada uno de los grupos.

En los casos en los que no se puede enmascarar, no queda más remedio que hacer un estudio abierto, aunque se puede intentar recurrir a un evaluador ciego. Esto quiere decir que, aunque investigadores y participantes conozcan la asignación al grupo de intervención o de placebo, la persona que analiza las variables desconoce esta asignación.

El enmascaramiento se puede hacer incluso cuando la vía de administración de los dos fármacos que se comparan sea distinta; esto se realiza administrando a cada grupo un placebo por la vía de administración distinta al fármaco que están recibiendo y se denomina estrategia de la doble simulación (doble *dummy*). Por ejemplo, el grupo del fármaco tópico recibe el fármaco por vía tópica y un placebo por vía intravenosa y el grupo del fármaco endovenoso recibe el fármaco por vía intravenosa y un placebo por vía tópica.

Tanto las reacciones adversas específicas como la falta de eficacia y los cambios analíticos específicos pueden desenmascarar un EC.

Otro aspecto que se tiene que considerar en esta fase del ensayo clínico es que siempre que exista un tratamiento eficaz, hay que administrarlo al grupo control, así que lo que se determinará con el EC será la eficacia relativa del nuevo fármaco. En caso de no existir alternativa terapéutica previa, se medirá la eficacia en términos absolutos.

■ Análisis de los resultados

En el análisis de los resultados (Tabla 5.2) se deben tener en cuenta los siguientes aspectos:

- Las **pérdidas de sujetos incluidos en el EC** ocurridas antes de la aleatorización van a afectar a la capacidad de generalización de los resultados, mientras que las pérdidas posaleatorización pueden afectar a la validez interna.
- El **análisis estadístico de los EC** es muy parecido al de los estudios de cohortes, pero es más frecuente el uso de métodos no paramétricos y, al tener importancia no solo que ocurra la variable-resultado, sino cuándo ocurre, es frecuente el empleo de análisis de supervivencia.
- **Comparaciones múltiples.** Al aumentar el número de comparaciones, se incrementa la posibilidad de resultados falsamente positivos ► MIR 17-18, 209.

- **Análisis de subgrupos.** Puede ocasionar problemas, especialmente cuando no se tiene previsto desde el principio. Produce comparaciones múltiples, aumenta la probabilidad de resultados espurios y, por tanto, sus resultados deben interpretarse con precaución.
- **Análisis por intención de tratar frente a análisis por protocolo:**
 - **Por protocolo.** Incluye únicamente a aquellos pacientes que han cumplido los requisitos del protocolo y/o han finalizado el estudio.
 - **Por intención de tratar.** El análisis incluye a todos los pacientes que han sido seleccionados y en el grupo en el que fueron asignados, aunque no hayan finalizado el estudio o hayan cambiado de grupo.
- Los **análisis intermedios** se realizan durante las fases III y IV del estudio. Están justificados para evitar que los pacientes del grupo control no se beneficien del tratamiento. Sin embargo, el realizar muchos análisis intermedios, aumenta el riesgo de cometer un error tipo I. Debe estar previsto en el protocolo del estudio, se ha de mantener la confidencialidad de los resultados y tiene que haber una penalización estadística porque, de nuevo, se están realizando comparaciones múltiples ► MIR 14-15, 193.
- La **magnitud del efecto del tratamiento** en los ensayos clínicos se puede expresar de varias formas:
 - **Riesgo relativo (RR).** Es el cociente entre el riesgo de sufrir un determinado evento en el grupo expuesto a un determinado tratamiento y el de sufrir el mismo evento en el grupo control (no expuesto al tratamiento).
 - **Reducción absoluta del riesgo (RAR).** Es la diferencia entre el porcentaje de eventos en el grupo control y el mismo porcentaje en el grupo experimental.
 - **Reducción relativa del riesgo (RRR).** Se define como la diferencia entre el porcentaje de eventos en el grupo control y el mismo porcentaje en el grupo experimental, dividido por el porcentaje de eventos en el grupo control.
 - **Número necesario de pacientes a tratar (NNT).** Se obtiene como el inverso de la RAR multiplicado por 100, y es el número de pacientes que se deben tratar para prevenir un suceso indeseable adicional.

PRAGMÁTICOS	Muestra heterogénea, validez externa
EXPLICATIVOS	Muestra homogénea, validez interna
CRUZADO	Cada grupo toma ambos fármacos con período de lavado
PARALELO	Cada grupo toma un único fármaco
POR INTENCIÓN DE TRATAR	Todos los sujetos incluidos son analizados
POR PROTOCOLO	Solo se analizan los que terminan el estudio
DE TAMAÑO PREDETERMINADO	Tamaño muestral conocido <i>a priori</i>
SECUENCIALES	Incorporación progresiva de pacientes al estudio
DE SUPERIORIDAD	Demostrar que un fármaco es mejor que otro
DE NO INFERIORIDAD	Se supone una eficacia similar, el nuevo fármaco tiene otras ventajas
ENMASCARADOS	Empleo de técnicas de ciego para evitar sesgos
ABIERTOS	No enmascarados

Tabla 5.2. Clasificación de los ensayos clínicos.



5.3. Tipos de ensayos clínicos

Los tipos de ensayos clínicos se exponen en los siguientes apartados.

■ Diseño clásico o en paralelo

El grupo control recibe el tratamiento a la vez que el grupo experimental, con el fin de controlar el efecto de factores pronósticos que pudieran cambiar a lo largo del tiempo.

El análisis de los datos consiste en comparar la diferencia en la variable-resultado entre ambos grupos con la variabilidad esperable dentro de cada grupo por el mero efecto del azar. Es muy apropiado en los casos de enfermedades agudas (Figura 5.2) ▶ MIR 16-17, 117; MIR 14-15, 196.



Figura 5.2. Ensayo clínico paralelo.

■ Diseño cruzado o intrapaciente

Consiste en que un mismo paciente va a recibir los dos tratamientos objeto de comparación en dos momentos distintos, de modo que el propio paciente sirve de control a sí mismo, lo que permite llevar a cabo el estudio con un tamaño muestral menor al del diseño clásico al reducir la variabilidad (Figura 5.3).

La principal ventaja del ensayo cruzado es eliminar la variabilidad *entre*-individuos en la respuesta a la intervención, pues estos ensayos valoran el efecto *intra*-individuo. Sus inconvenientes son:

- Al observar el efecto de un tratamiento, se asume la ausencia de efectos residuales del tratamiento, anterior (*carryover*). Debe existir un período de blanqueo o lavado suficientemente largo como para asegurar que no permanecen los efectos residuales del primer tratamiento asignado.
- La secuencia con la que se administran los diferentes tratamientos puede asociarse a respuestas diferentes (efecto período).
- Este diseño no es posible cuando la nueva terapia es quirúrgica o el primer tratamiento está dirigido a curar definitivamente la enfermedad (p. ej., en enfermedades agudas).
- Resulta difícil adjudicar los efectos secundarios tardíos.
- El análisis estadístico es más complejo que para grupos paralelos y requiere el uso de métodos apareados o de medidas repetidas.

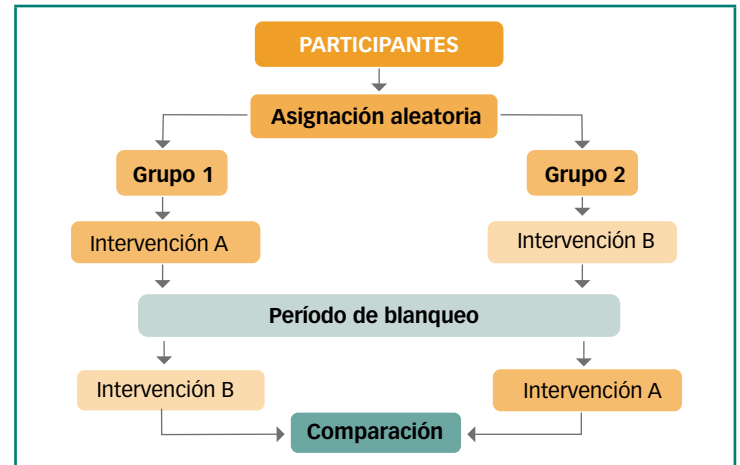


Figura 5.3. Ensayo clínico cruzado.

■ Diseño secuencial

Consiste en introducir pares de pacientes aleatorizados a los dos tratamientos hasta que la diferencia entre los distintos tratamientos favorezca a uno u otro (exceso de preferencias), momento en el que el ensayo clínico se detiene.

■ Diseño factorial

En ese tipo de diseño se evalúan simultáneamente dos tratamientos distintos en una misma muestra de sujetos, asignándose aleatoriamente los sujetos a uno de los cuatro grupos posibles: A, B, A + B y placebo (Tabla 5.3).

TRATAMIENTO	A	Placebo
B	A + B	B
Placebo	A	Placebo

Cada paciente se asigna a uno de los 4 grupos resultantes

Tabla 5.3. Ensayo clínico de diseño factorial.

En los ensayos factoriales, los placebos que se utilizan tienen que parecerse externamente al tratamiento activo asignado en cada secuencia. Es la razón por la que se habla de “placebo de A” y “placebo de B” para los dos tratamientos A y B, respectivamente. El diseño factorial habitualmente se utiliza cuando se quiere estudiar la eficacia de terapias combinadas. En este caso, los medicamentos utilizados pueden tener efectos diferentes cuando se combinan, y al final interesará comparar por separado los tres grupos con algún tratamiento activo frente al que solo recibe placebos (Figura 5.4) ▶ MIR 14-15, 178.

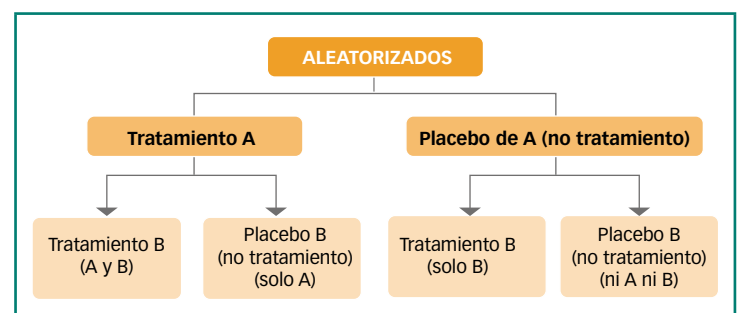


Figura 5.4. Ejemplo de asignación de tratamiento en el ensayo clínico factorial.

Diseño de no-inferioridad

Habitualmente los EC se realizan con el fin de demostrar la superioridad de un fármaco con respecto a otro. Sin embargo, puede ser útil probar que dos fármacos son al menos igualmente eficaces, con lo cual se podrá emplear, por ejemplo, el más barato. Es decir, importa que no sea inferior, de manera que si es igual o superior se considera que el fármaco experimental es "no-inferior" comparado con el grupo control ▶ MIR 17-18, 25.

Se denomina **equivalente terapéutico** a un fármaco diferente en su estructura química del original, pero del que se espera un efecto terapéutico y un perfil de efectos adversos similares cuando se administra a un paciente ▶ MIR 15-16, 29; MIR 15-16, 190; MIR 14-15, 190; MIR 13-14, 202.

Los ensayos de equivalencia valoran si dos terapias son iguales y se basan en este margen de mínima diferencia clínicamente importante (MCID o delta) que se fija *a priori*. Se considera que este es el límite a partir del cual una diferencia tendrá importancia desde el punto de vista práctico o clínico.

El objetivo es demostrar que la diferencia entre la nueva terapia y la convencional estará completamente dentro de ese margen. Se calculará un intervalo de confianza al 95% para la diferencia entre ambos grupos, y se demostrará la igualdad siempre y cuando ese intervalo de confianza de la diferencia entre tratamientos esté totalmente incluido dentro del margen delta, que suele ser muy pequeño. Debido a este planteamiento, los ensayos de equivalencia requieren mayor tamaño muestral, pues a mayor tamaño muestral más estrecho será el intervalo de confianza. Los ensayos de equivalencia intentan probar que una nueva terapia obtiene un resultado similar al tratamiento estándar, esto es, que el efecto del nuevo tratamiento está entre $-\text{delta}$ y $+\text{delta}$. Los ensayos de no-inferioridad intentan probar que la nueva terapia es igual, o si fuese peor que el tratamiento habitual, la diferencia entre ambos sería muy pequeña e inferior a delta.

Tal y como se ha comentado, la equivalencia terapéutica se define a partir de los estudios de no-inferioridad. En estos estudios es importante tener en cuenta dos aspectos (Figura 5.5):

- **Valor delta.** Es la máxima diferencia clínica que se acepta para definir dos tratamientos como equivalentes, que es distinta para cada tipo de fármaco analizado.
- **Análisis de las variables.** El análisis que suele realizarse es "por protocolo". Este análisis permite aumentar las diferencias entre los tratamientos, lo que dificulta poder concluir que dos tratamientos son equivalentes, manteniendo la posición más cauta en la interpretación de los resultados.

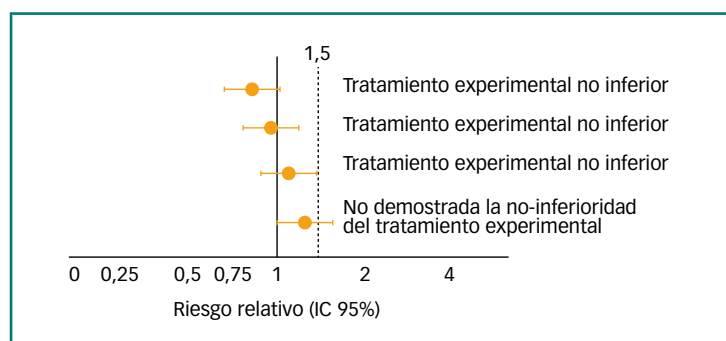


Figura 5.5. Interpretación de los resultados del estudio de no-inferioridad.

Estudios de bioequivalencia

Son EC en los que su objetivo es la demostración de la similitud en biodisponibilidad de dos formulaciones de un mismo principio activo a partir de la comparación de sus características farmacocinéticas. En la mayoría de los casos, el diseño de los estudios de bioequivalencia es el de un EC cruzado y con asignación aleatoria de dos secuencias de tratamiento.

Con el fin de reducir la variabilidad de la muestra, se suelen elegir para participar en el estudio voluntarios sanos (no persiguen, por tanto, comparar la eficacia de dos formas de tratamiento ▶ MIR 16-17, 36).

La demostración de bioequivalencia se obtiene mediante la comparación de los perfiles farmacocinéticos de los fármacos estudiados. Para ello, después de la administración de cada formulación, es necesario saber qué cantidad de fármaco existe en el organismo y cómo va variando a lo largo del tiempo. El procedimiento más habitual consiste en la obtención de sucesivas muestras de sangre. Es mucho menos frecuente que sea necesaria la determinación del fármaco en orina o en otras muestras biológicas.

Como medida de la cantidad de fármaco absorbido se utiliza el área bajo la curva concentración-tiempo (AUC, del inglés *area under curve*), y como indicador de la velocidad de absorción se mide la concentración máxima (C_{max}) alcanzada en la curva concentración-tiempo y el tiempo que tarda en alcanzarse (T_{max}). Se entiende por bioequivalencia entre dos productos cuando presentan una biodisponibilidad comparable en condiciones experimentales apropiadas.

De acuerdo con las normas de consenso (EMA, FDA), se considera que dos formulaciones son bioequivalentes cuando la diferencia en la velocidad y la magnitud de la absorción entre ellas es inferior al 20% (diferencias medias entre formulaciones comprendidas entre 0,8-1,2), en términos del intervalo de confianza (IC 90%) para la proporción entre las medias de las dos formulaciones comparadas (AUC test/AUC referencia y C_{max} test/ C_{max} referencia) (Figura 5.6 y Figura 5.7) ▶ MIR 15-16, 38.

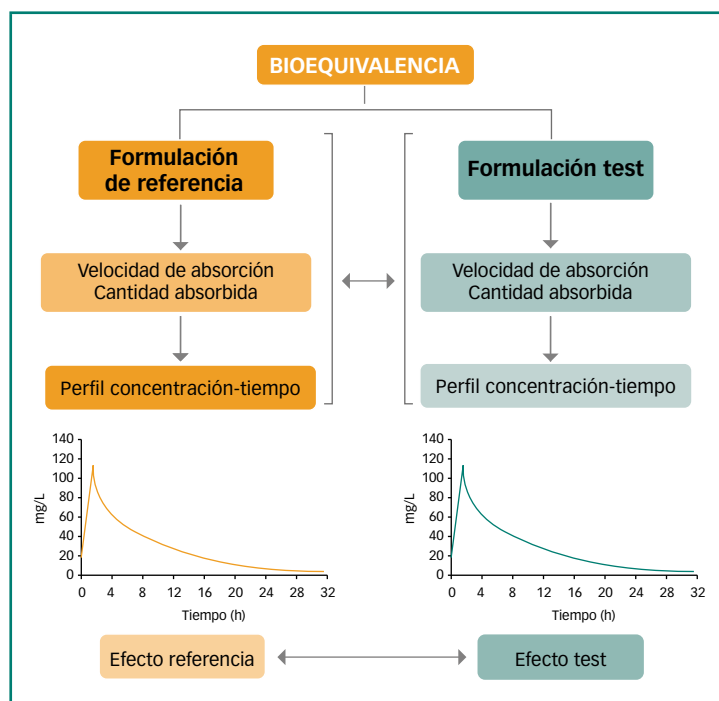


Figura 5.6. Principios para un estudio de bioequivalencia.

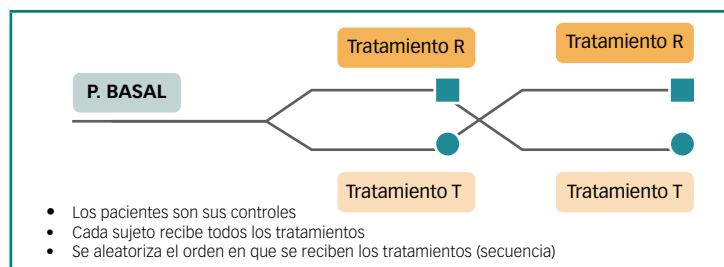


Figura 5.7. Diseño de un ensayo de bioequivalencia.

■ Ensayo clínico de tamaño único ($n:1$)

Se basa en modificar el tratamiento de una enfermedad de un paciente a lo largo de una serie determinada de períodos de tiempo. El fin del estudio es confirmar la eficacia o no del tratamiento en un único paciente.

Como en este caso hay un único paciente, no es posible compararlo con otros, por lo que aquí se comparan períodos de tratamiento para el mismo paciente. Por ejemplo, el paciente L.G. recibe durante un tiempo el tratamiento A y luego el B. Es decir, en un ensayo $n:1$, la unidad experimental es el período de tratamiento y no el paciente. Se compara la eficacia del tratamiento A frente al B en el señor L.G. Por supuesto, la secuencia de tratamientos puede hacerse al azar y con técnicas de enmascaramiento. El ensayo clínico $n:1$ es muy útil para:

- Evaluar el tratamiento sintomático en enfermedades en las que no se haya encontrado una intervención eficaz en ensayos convencionales como, por ejemplo, en enfermedades raras.
- En aquellas situaciones en las que exista una gran variabilidad interindividual en la respuesta al tratamiento, como en el caso del tratamiento con dicumarínicos.
- En tratamiento sintomático de enfermedades con importante componente subjetivo como, por ejemplo, migrañas, dolores articulares en la artrosis, fibromialgia...

Es necesario que la condición estudiada sea crónica, con síntomas repetidos pero reversibles y con respuesta al tratamiento evaluable por el paciente (cefalea tensional). Se evalúa, por tanto, un tratamiento sintomático y no curativo. Por las características del estudio es evidente que las conclusiones obtenidas no son generalizables a todos los pacientes con la misma enfermedad o síntoma.

5.4. Recomendaciones para la publicación de ensayos clínicos

Los estudios aleatorizados representan el patrón de referencia para conocer la eficacia de un determinado tratamiento. Sin embargo, la calidad de los estudios puede no ser óptima. La **declaración CONSORT** (*Consolidated Standards of Reporting Trials* [Normas consolidadas para las publicaciones de EC]) persigue mejorar la metodología de los ensayos clínicos, ya que permite presentar de forma sistematizada y con total transparencia los resultados para mejorar su validez.

Se basa en el seguimiento de una lista de variables en función de la evidencia (Tabla 5.4) y un diagrama de flujo (Figura 5.8) durante el proceso de escritura, revisión y análisis de los EC.

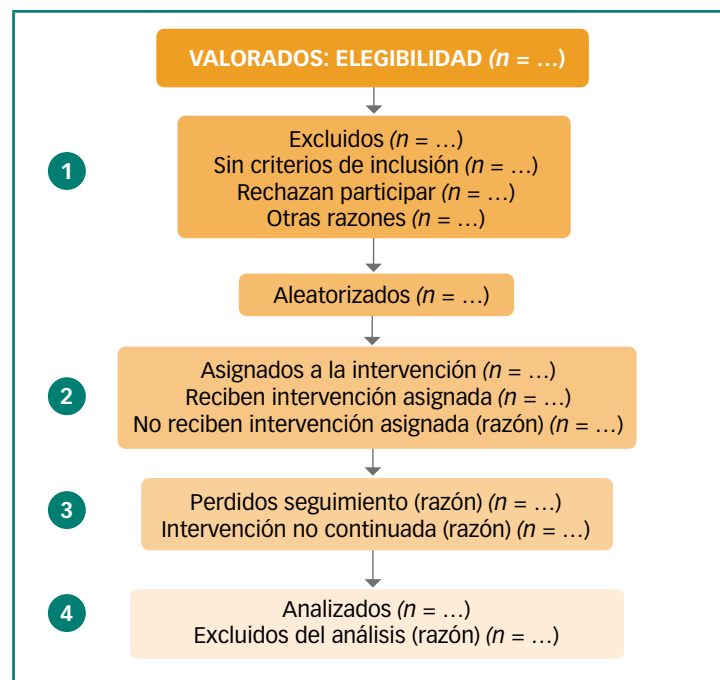


Figura 5.8. Diagrama de flujo de los sujetos participantes en las diferentes fases de un estudio aleatorizado: (1) período de inclusión, (2) asignación a la intervención, (3) seguimiento, (4) análisis.

Los puntos de la lista de comprobación conciernen al contenido del título, resumen, introducción, métodos, resultados y comentarios (discusión).

5.5. Agencia Española de Medicamentos y Productos Sanitarios

La Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), como agencia estatal adscrita al Ministerio de Sanidad, es responsable de garantizar a la sociedad, desde la perspectiva de servicio público, la calidad, seguridad, eficacia y correcta información de los medicamentos y productos sanitarios, desde su investigación hasta su utilización, en interés de la protección y promoción de la salud de las personas, de la sanidad animal y el medio ambiente.

Los comités científicos son órganos de asesoramiento que garantizan la transparencia e independencia en las actuaciones de la AEMPS. Son órganos de naturaleza mixta, con participación de responsables de la AEMPS y otros expertos de reconocido prestigio, así como representantes de las asociaciones de consumidores y usuarios o asociaciones profesionales de médicos, farmacéuticos o veterinarios. También existen comités de coordinación de la AEMPS con las autoridades de las comunidades autónomas y con los servicios periféricos de inspección farmacéutica.

Una de las fortalezas de la AEMPS es, sin duda, la interacción con los excelentes profesionales del Sistema Nacional de Salud y de las universidades españolas, a través de la Red de expertos de la AEMPS, que le presta asesoramiento científico y clínico en áreas específicas de conocimiento.

La AEMPS ha desarrollado la aplicación CIMA para realizar consultas bajo distintos criterios, para obtener distintos niveles de detalle de la información de los medicamentos. Esta información hace referencia a

la disponibilidad del medicamento en oficinas y/o servicios de farmacia, con su situación de autorización, estados de comercialización y posibles problemas de suministro. Todas las características de interés de un

medicamento se encuentran recopiladas en el detalle del mismo. Estas se presentan junto con el prospecto, ficha técnica y demás información necesaria en un medicamento ► **MIR 18-19, 132**.

VARIABLE (NÚMERO)		DESCRIPCIÓN
Título y resumen (1)		Cómo se asignaron a los pacientes a la intervención ("asignación aleatorizada", "aleatorizados" o "asignados aleatoriamente")
Introducción	Antecedentes (2)	Antecedentes científicos y explicación de la hipótesis
	Participantes (3)	Criterios de elegibilidad de los participantes, así como situación y lugar donde se obtuvieron los datos
	Intervenciones (4)	Detalles precisos de las intervenciones planeadas en cada grupo, y cómo y cuándo se realizaron finalmente
	Objetivos (5)	Objetivos e hipótesis específicos
	Resultados (6)	Definición clara de los objetivos de valoración de resultados primarios y secundarios y, en su caso, de los métodos utilizados para mejorar la calidad de las mediciones (p. ej., observaciones múltiples, entrenamiento de los valoradores)
	Tamaño de muestra (7)	Cómo se determinó el tamaño muestral y, en su caso, explicación de cualquier análisis interno y de las reglas para la finalización del estudio
	Aleatorización	Generación de la consecuencia (8)
		Ocultación de la asignación (9)
		Implementación (10)
	Cegamiento (enmascaramiento) (11)	
	Métodos estadísticos (12)	
Métodos	Flujo de participantes (13)	
	Selección (14)	
	Datos basales (15)	
	Números analizados (16)	
	Resultados y estimaciones (17)	
	Análisis secundarios (18)	
	Acontecimientos adversos (19)	
Resultados	Interpretación (20)	
	Generabilidad (21)	
	Evidencia global (22)	
Discusión	Interpretación de resultados en la que se tenga en cuenta la hipótesis del estudio, las posibles fuentes de sesgos o imprecisiones y los peligros asociados con la multiplicidad de análisis y de resultados	
	Generabilidad (validez externa) de los resultados del estudio	
	Interpretación general de los resultados en el contexto de la evidencia actualmente disponible	

Tabla 5.4. Lista de variables CONSORT.

Casos clínicos

Un equipo de investigadores realizó un estudio prospectivo para evaluar la eficacia de dos alternativas terapéuticas (técnica A frente a técnica B) en el tratamiento de las fracturas subcapitales de cadera. Los pacientes fueron asignados a uno u otro grupo de tratamiento de manera aleatoria. ¿De qué estudio se trata?

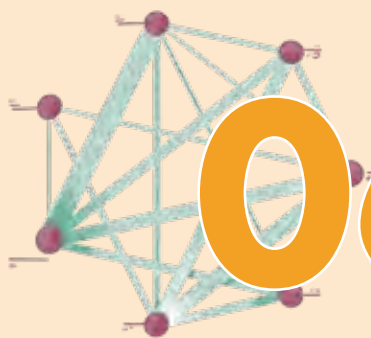
- 1) Cohortes.
- 2) Casos y controles.
- 3) Transversal.
- 4) Ensayo clínico.

RC: 4

En un ensayo clínico se evalúa la no-inferioridad de clopidogrel (experimental) frente a AAS (control). El límite clínicamente relevante es -5% en la incidencia de ictus isquémico en pacientes con estenosis carotídea grave. Los resultados mostraron una diferencia absoluta de eficacia de +1%; IC 95% (-8% a +12%). Señale la respuesta correcta:

- 1) Clopidogrel es superior a AAS.
- 2) AAS es no-inferior a clopidogrel.
- 3) El estudio no es concluyente.
- 4) Clopidogrel es no-inferior a AAS.

RC: 3



06 El metanálisis

Orientación MIR

Tema muy preguntado en los últimos años. Hay que conocer el concepto y diseño de un metanálisis, sus fases, y cómo y por qué se valoran la heterogeneidad de los estudios y la existencia de un posible sesgo de publicación.

Preguntas MIR

- ▶ MIR 22-23, 48
- ▶ MIR 21-22, 43
- ▶ MIR 20-21, 46
- ▶ MIR 17-18, 219
- ▶ MIR 16-17, 18
- ▶ MIR 15-16, 28
- ▶ MIR 14-15, 199; MIR 14-15, 200
- ▶ MIR 13-14, 194

Conceptos clave

- ⦿ El metanálisis es una técnica estadística que combina los resultados de estudios independientes, en general, ensayos clínicos aleatorizados.
- ⦿ Los metanálisis se utilizan, en general, para evaluar la efectividad de determinadas intervenciones-tratamientos, generando estimaciones más precisas que los estudios incluidos de forma individual y contando con mayor potencia estadística que ellos.
- ⦿ La validez de un metanálisis depende, entre otros, de la calidad de la revisión sistemática que lo acompaña.
- ⦿ Un buen metanálisis pretende incluir todos los resultados de los estudios relevantes sobre un tema, buscar la presencia de heterogeneidad en ellos y explorar la robustez de los hallazgos principales mediante un análisis de sensibilidad.

6.1. Introducción

El objetivo de una revisión sistemática es presentar un resumen imparcial y exhaustivo de las investigaciones (publicaciones) existentes, permitiendo la toma de decisiones acerca de una determinada intervención o tratamiento, basada en todos los estudios relevantes y de adecuada calidad existentes acerca del tema.

Algunas revisiones sistemáticas asocian además un análisis estadístico nuevo, combinando los diferentes estudios incluidos en la revisión: esto es un metanálisis ▶ MIR 13-14, 194.

Mucha de la evidencia de metanálisis y revisiones sistemáticas se obtiene de la red Cochrane, que es una red internacional de investigadores independientes que recopilan y aúnan en forma de revisión sistemática los mejores datos procedentes de la investigación biomédica para ayudar en la toma de decisiones ▶ MIR 17-18, 219.

6.2. Etapas del metanálisis

En la **Figura 6.1** se exponen las diferentes etapas que conforman un metanálisis.

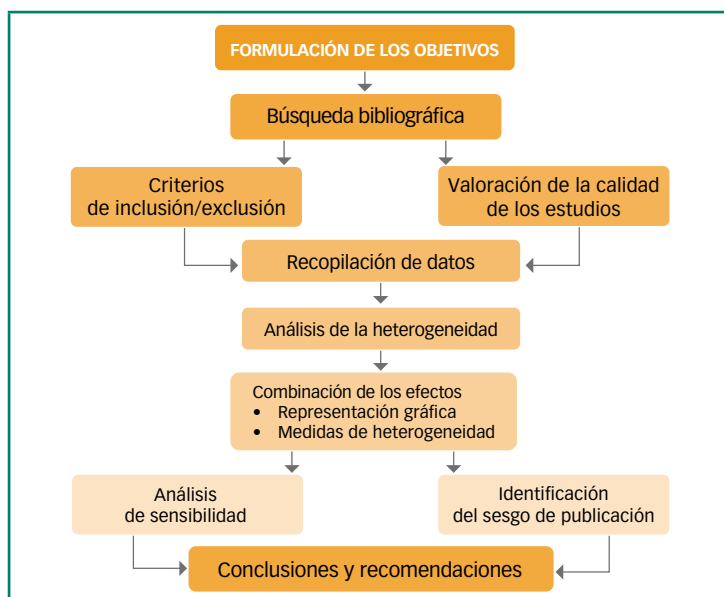


Figura 6.1. Seguimiento general de las etapas del metanálisis.

■ Análisis de heterogeneidad

Un **metanálisis** consiste en un análisis estadístico que combina o integra los resultados de varios ensayos clínicos independientes entre sí pero que el analista considera “combinables”. Una de las mayores dificultades y limitaciones consiste en decidir qué es o no combinable. Claramente resultarían combinables solo aquellos estudios en los que la hipótesis de investigación coincide por completo. Pero, desafortunadamente, los estudios difieren en los tipos de pacientes estudiados, la gravedad de la enfermedad, etc. Es por ello que todo metanálisis debe analizar la cantidad de heterogeneidad existente entre los estudios incluidos, esto es, debe contar con un análisis de heterogeneidad ▶ MIR 14-15, 199. Existen varios métodos estadísticos y gráficos para evaluar el grado de heterogeneidad.

Métodos estadísticos

Los métodos estadísticos pueden ser:

- **Prueba Q de Cochrane.** Es un estadístico basado en un test de ji al cuadrado que evalúa directamente la heterogeneidad entre los estudios incluidos. Una de sus limitaciones es su baja potencia, lo que hace, por un lado, que se utilice poco, y por otro, que se tienda a compensar considerándolo estadísticamente significativo a partir de valores de p por debajo de 0,10. Se podría resumir su interpretación de esta manera:
 - Si el valor de p correspondiente a la Q de Cochrane es estadísticamente significativo, existe alta probabilidad de heterogeneidad.
 - Si no es estadísticamente significativo, la interpretación se realiza haciendo un cociente entre el valor de la Q y sus grados de libertad. Si este cociente es menor de 1, la heterogeneidad es poco probable, si es mayor de 1, la heterogeneidad es posible.
- **Estadístico I².** Mucho más utilizado que el anterior, expresa el porcentaje de heterogeneidad existente en el metanálisis, con valores entre el 0% (no hay heterogeneidad, luego los estudios son homogéneos) y el 100% (el metanálisis es completamente heterogéneo). En general, se considera que valores por debajo del 25% indican baja heterogeneidad, y por encima del 75%, elevada.

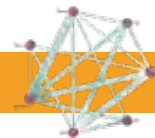
Métodos gráficos

El análisis visual del *forest plot* permite valorar rápidamente la presencia de heterogeneidad: si los intervalos de confianza de la medida de asociación empleada en los diferentes estudios se solapan entre sí, se puede considerar que el metanálisis incluye estudios homogéneos.

■ Combinación de los efectos

Según los resultados del análisis de heterogeneidad, la presencia o ausencia de esta determina el método de análisis ▶ MIR 21-22, 43; MIR 14-15, 200:

- Si no existe heterogeneidad o esta es muy baja, el análisis debe utilizar un modelo conocido como **modelo de efectos fijos**. Este modelo asume que existe un único efecto en la población, un efecto fijo poblacional que estará presente en todos los estudios, de manera que si existen resultados dispares entre ellos, será simplemente por azar. Así, el tamaño del estudio y su propia varianza (variabilidad intra-estudio) son los únicos determinantes del peso de cada estudio individual dentro del metanálisis.
- Si existe heterogeneidad, deberá emplearse el **modelo de efectos aleatorios**. Este modelo tiene en cuenta que los efectos de la exposición/intervención en la población son diversos y que los estudios incluidos en la revisión son solo una muestra aleatoria de todos los posibles efectos. Así, la ponderación de los estudios bajo este tipo de modelos considera no solo su propia varianza (variabilidad intra-estudio), sino también la que pueda existir entre los estudios (variabilidad entre-estudios). Este modelo tiene dos limitaciones fundamentales: la primera, tiende a aumentar la varianza del resultado del metanálisis, haciendo que la posibilidad de encontrar un resultado significativo disminuya (disminuye la potencia) y generando intervalos de confianza más amplios; la segunda, concede un peso excesivo a los estudios con pequeño tamaño muestral. Siempre que se detecte la presencia de heterogeneidad, debe investigarse su causa. Para ello se emplea una técnica denominada **metarregresión**. Esta técnica permite, por ejemplo, explorar si las características de los pacientes incluidos en los distintos estudios, o el diseño de los mismos, ha podido generar heterogeneidad.



Representación gráfica de los resultados (*forest plot* o diagrama de bosque)

La representación gráfica **MIR 20-21, 46** de los resultados de un metanálisis contribuye a una fácil y rápida interpretación. Para ello, se construye una gráfica en la que en el eje de abscisas (eje X) se representa la medida de efecto considerada (*odds ratio*, riesgo relativo...) y a lo largo del eje de coordenadas (eje Y) se sitúan los diferentes estudios, generalmente ordenados por el año de publicación o cualquier otro criterio de ordenación.

Para cada estudio y para la estimación global del efecto (resultado del metanálisis), se representa su estimación puntual y también el intervalo de confianza (IC) que le corresponde (asumiendo un modelo de efectos fijos y/o efectos aleatorios) **MIR 22-23, 48; MIR 15-16, 28**.

■ Identificación del sesgo de publicación

Si es importante detectar si existe o no heterogeneidad, también lo es determinar si el metanálisis ha tenido en cuenta todos los estudios importantes relativos a la hipótesis que se quiere analizar, y esto incluye tanto los estudios publicados como los no publicados.

En general, se sabe que existe una tendencia a no publicar los estudios con resultados negativos (sesgo de publicación).

Se han propuesto varios métodos para la exploración de la posibilidad y la cuantía del sesgo de publicación:

- El más simple consiste en realizar un **análisis de sensibilidad** y calcular entonces el número de estudios negativos realizados y no publicados que debería haber para modificar el sentido de una eventual conclusión "positiva", obtenida con un metanálisis. Si este número es muy elevado, se considera que la probabilidad de que el sesgo de publicación haya modificado sustancialmente los resultados es baja, y se acepta la existencia de las diferencias sugeridas por el metanálisis.
- También se puede examinar la posibilidad de sesgo de publicación con el método conocido como **gráfico en embudo** (*funnel plot*) (**Figura 6.2**). El *funnel plot* representa los resultados de los diferentes estudios incluidos en el metanálisis en un gráfico frente a su tamaño muestral. Se parte del supuesto de que los estudios que tendrían mayor probabilidad de no ser publicados serían los que no muestran diferencias (estudios

"negativos"), sobre todo si eran de pequeño tamaño. Inversamente, si hubiera sesgo de publicación, entre los estudios pequeños (en los que hay mayor probabilidad de que se alteren los resultados por azar), se tendería a publicar los que mostraran diferencias.

Desde un punto de vista visual, si el gráfico es simétrico y muestra la forma de un embudo invertido, se puede asumir que no existe sesgo de publicación. Si el gráfico es asimétrico, el metanálisis puede haber obviado los resultados de algunos estudios no publicados y existir sesgo de publicación **MIR 16-17, 18**.

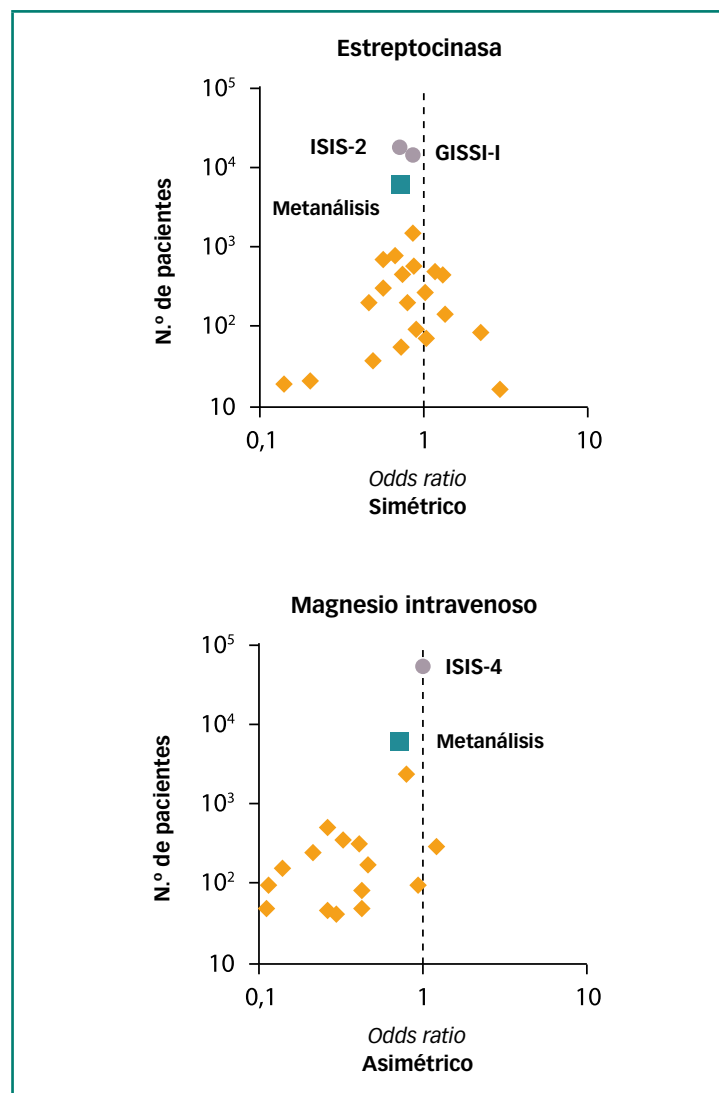
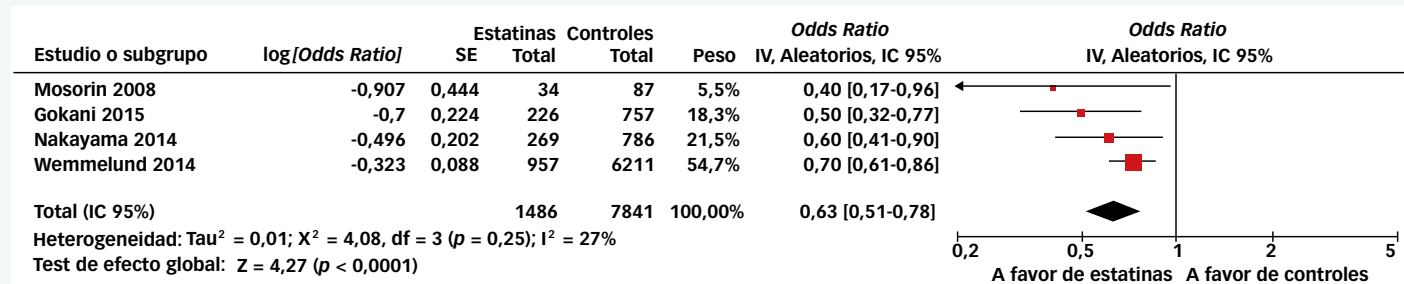


Figura 6.2. Gráfico en embudo o *funnel plot*: simétrico frente a asimétrico.

Casos clínicos

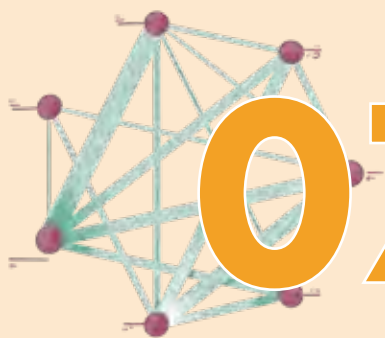
A continuación se muestran los resultados de un metanálisis.



A la vista de estos resultados, se puede concluir lo siguiente:

- 1) Presumiblemente se ha utilizado un modelo de efectos aleatorios, al contar con estudios muy heterogéneos.
- 2) El resultado del metanálisis no es estadísticamente significativo.
- 3) El estadístico I indica que existe un 0% de heterogeneidad.
- 4) Los estudios incluidos no pueden combinarse entre sí.

RC: 3



Medicina basada en la evidencia

Orientación MIR

Tema sencillo y de creciente importancia en el examen MIR.

En este tema vamos a presentar cómo se desarrolla el método científico, desde la realización de la pregunta de investigación hasta las recomendaciones de publicación.

Preguntas MIR

- ▶ MIR 22-23, 46
- ▶ MIR 20-21, 47; MIR 20-21, 53; MIR 20-21, 182
- ▶ MIR 19-20, 150
- ▶ MIR 16-17, 234

Conceptos clave

- ⦿ La mayor evidencia científica se obtiene de los diseños experimentales.
- ⦿ Se considera que un metanálisis de ensayos clínicos no sesgado es el estudio que aporta un mayor nivel de evidencia.
- ⦿ En cuanto a calidad del estudio el orden es (de mejor a peor) experimental, observacional, prospectivo, observacional retrospectivo descriptivo.

7.1. Niveles de calidad de la evidencia científica

■ Introducción

La **medicina basada en la evidencia (MBE)** se ha definido como la integración de la experiencia clínica individual con la mejor evidencia externa disponible procedente de la investigación sistemática (**Figura 7.1**).



Figura 7.1. Factores relacionados con la MBE.

El acto médico se entiende como una experiencia de relación interpersonal en la que los valores, las opiniones y la experiencia del médico, junto con las preferencias de los pacientes, tienen un papel preeminente, a lo que debe añadirse otro elemento crucial como es la evaluación sistemática de la evidencia científica. La experiencia clínica, entendida como el dominio creciente del conocimiento y el juicio que cada médico adquiere mediante la práctica profesional, se refleja, sobre todo, en su capacidad para alcanzar un diagnóstico preciso y para identificar e integrar los problemas, circunstancias y preferencias de cada paciente.

La evidencia externa procede de las ciencias básicas y de la investigación clínica, y establece la precisión de las pruebas diagnósticas, el poder pronóstico de los marcadores de riesgo, así como la eficacia y la seguridad de las intervenciones terapéuticas, rehabilitadoras o preventivas. Ambos aspectos son fundamentales y se entrelazan para alcanzar una elevada calidad asistencial. Se complementan con el respeto de la autonomía del paciente, incluyendo explícitamente los valores y las preferencias de este y de la sociedad en el proceso de toma de decisiones clínicas.

■ Estilo de vida basado en la evidencia

El panel de expertos voluntarios en materia de medicina preventiva de Estados Unidos (*US Preventive Services Task Force, USPSTF*) recomienda la adopción de la estrategia de prevención de las cinco aes (averiguar, aconsejar, acordar, ayudar y asegurar) para la ejecución de intervenciones conductuales y de consejo sobre los principales factores de riesgo, basándose en revisiones de la evidencia científica ► **MIR 20-21, 53**.

■ Factores determinantes de la calidad de la evidencia científica

Los aspectos del diseño de un estudio epidemiológico que están asociados a una mayor calidad y rigor científico son los siguientes:

- El sentido prospectivo del estudio (secuencia temporal).
- La asignación aleatoria a los grupos experimental y control.
- La existencia de un grupo control concurrente.
- El enmascaramiento de pacientes e investigadores participantes.
- La inclusión en el estudio de un número suficiente de pacientes como para detectar, en caso de haberlas, diferencias estadísticamente significativas entre los distintos grupos del estudio.

■ Escalas de gradación de la evidencia científica

Todas las escalas de gradación son muy similares entre sí y presentan al metanálisis de ensayos clínicos controlados y aleatorizados como el estudio más riguroso y que aporta mayor evidencia causal. La medicina basada en la evidencia persigue la integración de la maestría clínica individual con las mejores evidencias científicas disponibles (**Tabla 7.1** y **Tabla 7.2**)

► **MIR 20-21, 182; MIR 16-17, 234**.

ASSES-AVERIGUAR	Preguntar sobre los factores y las conductas de riesgo, así como acerca de los aspectos que afectan a la elección o el cambio de la conducta.
ADVISE-ACONSEJAR	Dar consejo claros, específicos y personalizados, e incluir información sobre los riesgos/beneficios personales.
AGREE-ACORDAR	Pactar colaborativamente los objetivos y los métodos más apropiados, basados en los intereses y en la capacidad para el cambio de la persona.
ASSIST-AYUDAR	Usar técnicas de modificación de la conducta (autoayuda o asesoramiento) para ayudar a la persona a conseguir los objetivos pactados adquiriendo las habilidades, la confianza y el apoyo social/ambiental que favorece el cambio, junto con los tratamientos farmacológicos cuando sean adecuados.
ARRANGE-ASEGURAR	Fijar (asegurar) visitas de seguimiento (en el centro o telefónicas) para ayudar/apoyar y para ajustar el plan terapéutico como se necesite, inclida la derivación a unidades especializadas cuando sea necesario.

Tabla 7.1. Las 5 aes de las estrategias de prevención.

A	Datos procedentes de múltiples ensayos clínicos aleatorizados o metanálisis.
B	Datos procedentes de un único ensayo clínico aleatorizado o de grandes estudios no aleatorizados.
C	Consenso de opinión de expertos y/o pequeños estudios, estudios retrospectivos o registros.

Tabla 7.2. Niveles de evidencia.

La **calidad de la evidencia** científica se clasifica en **tres grados** de recomendación para la adopción de una medida sanitaria:

- **Buena (grado A).** Existe buena evidencia científica para recomendar o desaconsejar la adopción del procedimiento médico. Son datos derivados de estudios aleatorizados y metanálisis.
- **Regular (grado B).** Hay evidencia científica para recomendar o desaconsejar la adopción del procedimiento médico, basada en EC aislados y diseños observacionales robustos.



- **Mala (grado C).** Existe evidencia científica para recomendar o desaconsejar la adopción del procedimiento médico basada en diseños de menor calidad metodológica.

Basados en los niveles de evidencia, se pueden realizar **recomendaciones científicas que aconsejan o desaconsejan las medidas analizadas**, como muestra la **Tabla 7.3**.

CLASE	DEFINICIONES	
I	Evidencia y/o acuerdo general de que un determinado procedimiento diagnóstico/tratamiento es beneficioso, útil y efectivo	Se recomienda/ está indicado
II	Evidencia conflictiva y/o divergencia de opinión sobre la utilidad/eficacia del tratamiento	
IIa	El peso de la evidencia/opinión está a favor de la utilidad/eficacia	Se debe considerar
IIb	La utilidad/eficacia está menos establecida por la evidencia/opinión	Se puede considerar
III	Evidencia o acuerdo general de que el tratamiento no es útil/efectivo y en algunos casos puede ser perjudicial	No se recomienda

Tabla 7.3. Recomendaciones.

Transferencia de la evidencia. Las guías de práctica clínica **MIR 22-23, 46; MIR 20-21, 47** son recomendaciones desarrolladas de forma sistemática para ayudar a profesionales y pacientes a tomar decisiones sobre la atención sanitaria más apropiada y a seleccionar las opciones diagnósticas o terapéuticas más adecuadas a la hora de abordar un problema de salud o una condición clínica específica. La importancia y necesidad de la realización de estas guías viene dada por los siguientes factores:

- Variabilidad en la práctica clínica por la existencia de áreas de incertidumbre.
- Existencia de un problema de salud importante con impacto en la morbilidad/mortalidad.
- Aparición de técnicas o tratamientos novedosos.
- Posibilidad de conseguir un cambio para mejorar resultados en la atención porque el proceso es susceptible de mejorarse por una actuación sanitaria y los medios para lograrlo están disponibles.
- Área de prioridad en el Sistema Nacional de Salud.

Para realizar una guía deben seguirse las fases expuestas en la **Figura 7.2**:

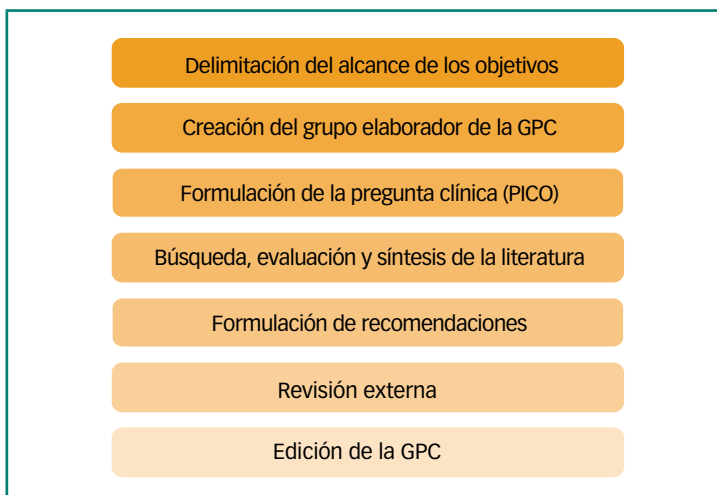


Figura 7.2. Fases de creación de una GPC.

El sistema GRADE

La MBE facilita moverse en este terreno incierto y ayuda en la toma de decisiones clínicas. En los últimos años, el sistema GRADE (por su acrónimo en inglés: *Grade of Recommendation, Assessment, Development, and Evaluation*) se ha utilizado en las publicaciones de guías o recomendaciones, principalmente de tratamiento, tanto farmacológicas como no farmacológicas en diferentes áreas de la medicina.

El grupo de trabajo GRADE es una colaboración internacional multidisciplinaria de expertos que desarrollan un sistema común, transparente y razonable para calificar la calidad de la evidencia y la fuerza de las recomendaciones.

En el desarrollo de GRADE, los autores consideraron un rango amplio de preguntas clínicas, prefiriéndose el método PICO (véase el subtema 9.1). Al formular las preguntas clínicas en formato PICO, queda definido de forma concreta y sin ambigüedad lo que se pretende averiguar.

El sistema GRADE propone las siguientes definiciones:

1. Calidad de la evidencia (certidumbre): indica hasta qué punto la estimación de un efecto es adecuada para apoyar una recomendación.
2. Fuerza de la recomendación: indica si poner en práctica la recomendación conllevará más beneficios que riesgos.

Calidad de la evidencia

Aunque la calidad de la evidencia es un espectro continuo, GRADE propone una clasificación en cuatro categorías (**Tabla 7.4**):

ALTA	La confianza del estimador del efecto disponible en la literatura científica es alta; se encuentra muy cercano al efecto real
MODERADA	Es probable que el estimador del efecto se encuentre cercano al efecto real, aunque podrían existir diferencias sustanciales
BAJA	El estimador del efecto puede ser sustancialmente diferente al efecto real
MUY BAJA	Es muy probable que el estimador del efecto sea sustancialmente diferente al efecto real

Tabla 7.4. Calidad de la evidencia según GRADE.

La calidad de la evidencia para un determinado evento comienza con el diseño de los estudios (calidad alta para ensayos clínicos y baja para estudios observacionales). A partir de ellos, aplica los 8 criterios para bajar o subir la calidad de la evidencia, 5 de los 8 criterios tienen el potencial de disminuir la calidad de la evidencia:

1. **Riesgo de sesgos:** posibilidad de limitada validez interna. Por ejemplo, ausencia de enmascaramiento o aleatorización, diseño por protocolo, etc.
2. **Inconsistencia de los resultados a través de los estudios:** significa que los resultados se desvían unos con otros, y esto naturalmente conduce a una confianza disminuida de la efectividad de la estimación.
3. **Ausencia de evidencia directa:** acontece cuando la variable de medida es una deducción a partir de estudios indirectos. Por ejemplo, si A es mejor que B y B es mejor que C, concluimos que C es mejor que A. Sin embargo, esta conclusión puede no ser necesariamente correcta.
4. **Imprecisión o variación aleatoria en la estimación del evento final:** error aleatorio que condiciona la amplitud de IC95%, acercándose al valor nulo de la estimación. Por ejemplo, un RR para una quimioterapia experimental vs. control de 0,7 IC95% (0,3-0,9) es significativa.

Sin embargo, si la quimio es muy tóxica un valor de 0,9, próximo al nulo, puede no aconsejar el uso del fármaco experimental dada la relación beneficio-riesgo.

5. **Sesgos de publicación:** véase los subapartados de metanálisis, análisis de sensibilidad y diagrama de embudo.

Se han propuesto 3 criterios más con el potencial de incrementar la confianza:

1. Una **fuerte asociación**, bien protectora o de riesgo, es más improbable que se deba a una sobrestimación por factores de confusión y es más probable que el efecto observado sea más cercano al efecto real.
2. Existencia de **gradiente dosis-respuesta**: es útil para establecer con mayor certidumbre una relación causa-efecto.
3. Evidencia de que todos los posibles **factores de confusión o sesgos** que podrían haber reducido el efecto están controlados.

Fuerza de la recomendación

La fuerza de una recomendación refleja el grado de certeza de que los efectos deseables de una intervención recomendada superan sus efectos no deseables. El sistema GRADE considera 2 categorías en relación con la fuerza de las recomendaciones:

- **Recomendación fuerte:** existe un claro desequilibrio riesgo-beneficio, bien a favor de los riesgos o de los beneficios.
- **Recomendación débil:** los efectos deseables probablemente superan los efectos no deseables o los efectos no deseables probablemente son mayores que los efectos deseables pero con una incertidumbre apreciable.

7.2. El método científico

■ La pregunta inicial. El método PICO

Cuando planteamos la realización de un estudio epidemiológico, es clave saber qué es lo que queremos medir. Esta pregunta suele enmarcarse en lo que se denomina PICO (de las iniciales en inglés *Patient, Intervention, Comparison and Outcome*). Aunque los componentes intervención y comparación sugieren el carácter experimental del ensayo clínico y parecen limitar el sistema a las cuestiones sobre tratamiento, puede utilizarse igualmente para preguntas sobre etiología, diagnóstico o pronóstico, como resume, a modo de ejemplo, la [Tabla 7.5](#) ▶ [MIR 19-20, 150](#).

Lo que plantea este sistema de formulación de preguntas es la necesidad de que el investigador tenga ordenada la necesidad de la puesta en marcha de dichas cuestiones y los estudios convenientes, de forma habitual, para dar respuesta a dichas preguntas ([Tabla 7.6](#)).

OBJETIVO PRINCIPAL (EJEMPLO)	TIPO DE ESTUDIO ACONSEJADO
Eficacia de una intervención (reducción de la carga viral con dos pautas diferentes de tratamiento)	Ensayo clínico aleatorio
Validación de una prueba diagnóstica (eficacia de la mamografía como test de <i>screening</i> o cribado)	Diseño transversal
Evaluación pronóstica (incidencia de reagudización de EPOC según dejen o no de fumar los pacientes)	Cohortes
Evaluación causal (asociación entre el tabaquismo y la cardiopatía isquémica, Framingham)	Cohortes (mejor), casos y controles

Tabla 7.6. Estudio aconsejado para dar respuesta a la pregunta epidemiológica (PICO).

Por ejemplo, aunque el mejor estudio en cuanto a aportar evidencia es el ensayo clínico, difícilmente se podrá hacer este estudio para evaluar causalidad.

■ Fuentes de información

Son el conjunto de actividades encaminadas a localizar y recuperar documentos incluidos en fuentes de información. Estas pueden ser:

- **Fuentes primarias.** Son libros y revistas que contienen estudios originales (ensayos clínicos, casos-control, etc.).
- **Fuentes secundarias.** Son las bases de datos, repositorios, índices, etc. que nos presentan la información ordenada y dirigida.
- **Fuentes terciarias.** Ofrecen información que ha sido revisada y valorada por expertos. Lo forman las bases de datos de evidencia como la Biblioteca Cochrane, Guía Salud, NICE, Epistemonikos y otros elementos, como las Guías de la RNAO.

Estas búsquedas se realizan en bibliotecas (fuentes primarias), en **bases de datos**, definidas como el conjunto de referencias bibliográficas de publicaciones almacenadas en formato electrónico, y que pueden ser recuperadas interactivamente gracias a un lenguaje específico de consulta y en **metabuscadores**, sistema que realiza la búsqueda de forma simultánea en varias bases de datos y presenta los resultados agrupados, destacando:

- **Medline.** Proporciona ayuda para búsquedas de literatura sanitaria, pero para obtener respuestas a problemas clínicos importantes puede ser menos útil. Actualmente, es la mejor fuente que se conoce para buscar los datos necesarios para resolver problemas en los distintos campos de ciencias de la salud.

	PACIENTE	INTERVENCIÓN	COMPARACIÓN	RESULTADO (OUTCOME)
Etiología	¿Se asocia el tabaquismo a la presencia de aneurismas de aorta abdominal (AAA)?			
	Adulto no fumador	Pacientes fumadores	Pacientes no fumadores	Prevalencia de AAA
Diagnóstico	¿Cuál es la validez del dímero-D en el diagnóstico inicial de TVP?			
	Pacientes con TVP	Determinación de dímero-D	Comparación con eco-doppler	Sensibilidad y especificidad del dímero-D
Pronóstico	¿Se puede evaluar el riesgo de rotura de los AAA?			
	Pacientes con AAA	AAA > 5,5 cm	AAA < de 5,5 cm	Riesgo de rotura según el diámetro del AAA
Tratamiento	¿Cuál es el mejor tratamiento para la estenosis carotídea grave sintomática?			
	Paciente con estenosis > 70%	Endarterectomía carotídea	Stent carotídeo	Incidencia de ictus posprocedimiento

Tabla 7.5. Método PICO para la puesta en marcha de preguntas en epidemiología.



- **SciELO (Scientific Electronic Library Online).** Es una biblioteca electrónica principalmente utilizada en países de habla hispana.
- **Base de datos de la Colaboración Cochrane.** Es una red internacional cuyo sentido es disponer, agrupar, sostener y divulgar revisiones sistemáticas y metanálisis de todos los ensayos clínicos controlados aleatorizados sobre los efectos de las intervenciones sanitarias en humanos que sirvan como ayuda en la práctica clínica, así como crear directrices para la toma de decisiones de los gestores de la salud y orientar investigaciones futuras.
- **PubMed.** Servicio gratuito de búsqueda de la US National Library of Medicine que permite el acceso a Medline y Pre-Medline (con enlaces a las revistas en línea incluidas), así como a otras bases de datos de referencia. El sistema de recuperación de la información enlaza con artículos a texto completo de unas 400 revistas (algunas requieren suscripción previa) y proporciona, asimismo, acceso a artículos relevantes que todavía no se han incorporado a Medline. Como tal, no es una base de datos, si no el buscador que nos da acceso a ciertas bases de datos de referencia.
- **Web of Science (WOS).** Facilita el acceso a un conjunto de bases de datos en las que aparecen citas de artículos de revistas científicas, libros y otros tipos de material impreso que abarcan todos los campos del conocimiento académico.

■ Cómo hacer la búsqueda: tesauros. Marcadores booleanos

Para realizar una búsqueda bibliográfica debemos conocer las herramientas de las que disponemos para realizar una búsqueda eficiente. Estas son:

- **Tesauros.** Vocabularios especializados, controlados y estructurados formalmente mediante relaciones de jerarquía, equivalencia y asociación, que actúan como sistemas de recuperación de información en las bases de datos.
 - **MeSH (Medical Subject Heading).** Utilizado por Medline y otras bases de datos biomédicas.
 - **DeCS (Descriptores en Ciencias de la Salud).** Creado por BIREME (acrónimo de Biblioteca Regional de Medicina, actualmente Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud), utilizado por LILACS y Medline, entre otras. Es un vocabulario estructurado, dinámico y trilingüe (español, inglés y portugués) con una estructura jerárquica.
- **Operadores booleanos.** Permiten combinar varios términos en una búsqueda y amplían o reducen el resultado de esta. Son los siguientes:
 - **AND (Y).** Indica que las palabras que anteceden y siguen al operador deben encontrarse en el resultado de la búsqueda. Ejemplo: *teenage and depression* arrojará resultados que contengan las palabras "joven" y "depresión".
 - **NOT (No).** Indica que la palabra clave anterior al operador deberá aparecer, pero no la posterior. Ejemplo: *depression not teenage* mostrará los resultados en los que aparezca "depresión" y no aparezca "joven" en los mismos.
 - **OR (O).** Indica que alcanza con que tan solo una de las palabras esté presente. En la mayoría de las herramientas de búsqueda puede reemplazarse por un espacio en blanco. Ejemplo: *teenage or depression* dará resultados que contengan cualquiera de las palabras.

- **Operadores de truncamiento y comodines.** Símbolos que sustituyen uno o varios caracteres de una palabra, de forma que permiten localizar un término y sus posibles variantes. El asterisco (*) es un signo que reemplaza un carácter o grupo de caracteres, y resulta muy útil cuando se desconoce el modo completo de la escritura del término. Se utiliza el asterisco cuando se omiten una o varias letras y (?) cuando se omite una sola, tanto si es en el medio o al final de la palabra.
- **Limitadores/Filtros.** Acotan la búsqueda por tipo de estudio, idioma, edades, sexo, subtemas, etc.

■ Indicadores bibliométricos

En la publicación de los resultados es importante elegir la revista que más se adecue al tema y tener en cuenta las características de esta. Para ello, se puede recurrir al **Journal Citation Report (JCR)** o la **Scimago Journal Rank (SJR)**, que ordenan las revistas según el índice de impacto que presentan.

Los parámetros que tienen en cuenta estos *ranking* son:

- **Factor de impacto.** Identifica la frecuencia con que se cita un artículo promedio de una revista en un año determinado. Se utiliza esta cifra para evaluar y comparar la importancia relativa de una revista con otras del mismo campo, o para ver con qué frecuencia se citan los artículos y determinar qué revistas pueden ser mejores las mejores en su área. Por ejemplo:

$$FI = \frac{\text{(citas hechas durante el 2020 a los artículos de 2019 y 2018)}}{\text{(número de artículos publicados en los años 2019 y 2018)}}$$

- **Índices de Eigenfactor.** Tiene en cuenta no solo los recuentos de las citas que recibe una revista, sino también toda la estructura de la red de citas para medir la influencia de estas en la literatura especializada. Esta métrica está disponible para JCR desde el año 2007 en adelante.
- **Índice H.** H es el número de artículos H que tienen al menos H citaciones. Por ejemplo: un índice H de 5 significa que hay 5 artículos que han sido citados 5 veces o más. Esta métrica es útil porque descuenta el peso desproporcionado de los artículos más citados o de los artículos que aún no han sido citados.

■ Normas de publicación

Los estándares de publicación son guías que publican las revistas científicas para ayudar a los autores a redactar el artículo final, aunque no están desarrollados para este fin, también pueden utilizarse en el proceso de lectura crítica.

Destacamos los siguientes:

- Declaración CONSORT (ensayos clínicos aleatorizados).
- Declaración TREND (ensayos clínicos no aleatorizados).
- Declaración STROBE (estudios observacionales).
- Declaración STARD (estudios de diagnóstico/pronóstico).
- Criterios QUORUM-PRISMA. PRISMA actualiza los criterios QUORUM para revisiones sistemáticas o metanálisis.
- Criterios COREQ. Estudios cualitativos.
- Criterios CARE. Informes de caso.

■ Estructura de un artículo científico

El artículo científico es un documento escrito y publicado que describe los resultados originales de la investigación. Su estructura habitual contiene: introducción, metodología, resultados y discusión, que se resumen con el acrónimo IMRyD (Figura 7.3).

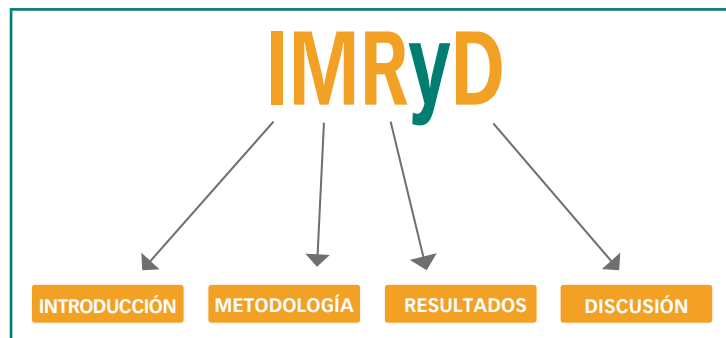


Figura 7.3. IMRyD.

- **Introducción.** Incluye una breve revisión del tema. Debe estructurarse en antecedentes, estado actual del tema y justificación. La redacción debe hacerse en tiempo presente, ya que se está plasmando la situación actual del tema y referenciando todos los autores.
- **Material y métodos.** En este apartado se describe el diseño de investigación y cómo se ha llevado a cabo. Además, debe hacerse de tal forma que, si el lector u otros investigadores quisieran hacer una réplica del estudio, pudieran realizarla con los datos facilitados. Es la parte más importante y más amplia del artículo.
- **Resultados.** Deben seguir una sucesión lógica de acuerdo con la estrategia de análisis empleada: datos descriptivos, resultados principales, análisis por grupos e informar sobre las pérdidas y los motivos por los que se han producido. La información no debe duplicarse, si aparece en tablas o figuras, no debe desarrollarse en texto.
- **Discusión.** Consiste en la interpretación de los resultados comparándolos con los encontrados en otros estudios. Debe incluir una evaluación de la validez interna del estudio comentando las limitaciones y las fuentes de error, así como de la validez externa, es decir, el grado en que pueden extrapolarse los resultados exponiendo sus implicaciones teóricas o prácticas, con prudencia e imparcialidad. La exposición finaliza indicando las futuras líneas de investigación y se especifican qué aspectos concretos deberían investigar.

- **Bibliografía.** En ella se incluyen los artículos mencionados en el texto. Estos artículos pueden enumerarse siguiendo distintas normas de citación como las de APA, Vancouver o ISO.

Las citas a referencias electrónicas son cada vez más abundantes, pero poco estables, ya que pueden cambiar de web o desaparecer. Por ello, se ha hecho necesario generar métodos de información permanente de los recursos electrónicos. Así surge en 2001, el *Digital Object Identifier* (DOI), un código alfanumérico que identifica en la web un artículo y lo recupera incluso si este se ubica en un servidor distinto al que fue alojado en un principio. En 2011, se hizo obligatorio para todas las revistas.

El DOI es, por tanto, un enlace permanente en forma de código alfanumérico que identifica de forma única un contenido electrónico. Esta pieza específica de contenido intelectual puede ser un artículo científico, una imagen, un libro, una canción u otro, siempre que se trate de un objeto en el ambiente digital. A diferencia del sistema de las páginas web, el sistema DOI no cambia con el paso del tiempo, aunque el artículo sea reubicado en una dirección distinta.

■ Citas y referencias bibliográficas. Gestores bibliográficos

En la elaboración de un artículo médico es muy importante el orden en el que se presenta la bibliografía en la que se basa dicho artículo. Para ello nos ayudamos de las citas y las referencias bibliográficas.

Las citas se pueden referir en el texto de diferentes formas, siendo las más habituales a modo de superíndice o entre paréntesis. Las utilizamos cuando nos referimos a un artículo previo publicado sobre el tema que estamos investigando. El artículo al que se refiere dicha cita se expresa como listado incluyendo autores, revista y título del artículo. Cada una de las entradas de este listado es lo que se denomina **referencia bibliográfica**. Esas referencias bibliográficas podrán seguir distintas normas como la APA o Vancouver.

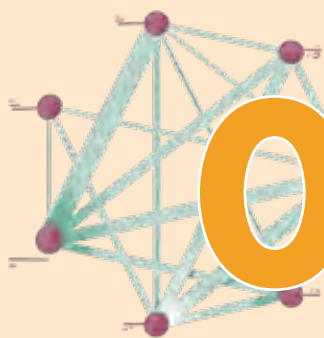
Los gestores bibliográficos son softwares que permiten almacenar, organizar, compartir e insertar bibliografía en diferentes formatos. Los más utilizados son **RefWorks**, **EndNote** (asociado a la WOS), **Mendeley** (además red social académica) y **Zotero**.

Casos clínicos

Indique cuál de los siguientes diseños epidemiológicos se asocia con una peor calidad en la relación causa-efecto entre una exposición y un evento:

- 1) Ensayo clínico.
- 2) Estudio de casos y controles anidado.
- 3) Ensayo de campo.
- 4) Estudio de tendencias temporales.

RC: 4



08 Estadística. Generalidades

Orientación MIR

En este tema solo es necesario saber dos cuestiones: los tipos de muestreo y los tipos de variables.

Preguntas MIR

- ▷ MIR 16-17, 130
- ▷ MIR 14-15, 184
- ▷ MIR 13-14, 190

Conceptos clave

- ⊙ La obtención de una muestra a partir de la población puede hacerse de forma probabilística (mediada por el azar) y no probabilística.
- ⊙ Los tipos de muestreo probabilístico son el aleatorio simple, el aleatorio estratificado, el sistemático y por grupos o conglomerados.
- ⊙ El estratificado debe realizarse cuando haya en la muestra una hipotética variable pronóstica que pudiera influir en el desarrollo de la enfermedad.
- ⊙ Las variables son características observables. Pueden ser cuantitativas o cualitativas, en función de que la variable analizada pueda cuantificarse o expresarse en función de una característica.
- ⊙ La forma de representación gráfica va a depender del tipo de variable. Para las cualitativas y las cuantitativas discretas, se emplea el diagrama sectorial y el diagrama de barras. Para las cuantitativas continuas, se utiliza el histograma y el polígono de frecuencias.
- ⊙ Para definir correctamente una variable en una muestra, se necesitará un parámetro que sirva para aglutinar todos los valores de dicha muestra (parámetro de tendencia central) y otro que sirva para informar de lo agregados o no que están los valores (parámetro de dispersión).
- ⊙ El tipo de distribución más frecuente es la normal o de Gauss, que es simétrica. En distribuciones simétricas se usará como parámetro de tendencia central la media, y de dispersión, la desviación típica.
- ⊙ En distribuciones asimétricas, se emplea la mediana como parámetro de tendencia central y el rango como parámetro de dispersión.
- ⊙ En la distribución de Gauss, media, mediana y moda coinciden en el mismo valor.
- ⊙ El coeficiente de variación es el parámetro que se emplea para decir si una distribución es homogénea o dispersa. Además, informa de cuál de entre dos distribuciones tiene una mayor variabilidad (dispersión).
- ⊙ La asimetría de una distribución mide la desviación "horizontal" desde la curva normal.
- ⊙ La curtosis mide el apuntamiento de una distribución.

Se define la **estadística** como el método de razonamiento que permite interpretar un conjunto de datos cuyo carácter esencial es la variabilidad.

8.1. Tipos de estadística

Existen dos tipos:

- **Estadística descriptiva.** Organización, presentación y síntesis de los datos de una manera científica y lógica.
- **Estadística inferencial.** Bases lógicas mediante las cuales se establecen conclusiones relacionadas con poblaciones, a partir de los resultados obtenidos en muestras. Trabaja con los datos que le proporciona la estadística descriptiva y es la parte de la estadística esencial para la medicina clínica. Se fundamenta en dos pilares básicos, que son el contraste de hipótesis y la estimación de intervalos de confianza.

8.2. Población y muestra

La **población** es el conjunto de todos los individuos, generalmente inaccesible, que tienen en común alguna característica observable y del que se pretende obtener una serie de conclusiones.

Se denomina **muestra** al conjunto menor de individuos, accesible y limitado, sobre el que se realiza el estudio con idea de obtener conclusiones generalizables a la población.

El proceso de selección de los individuos se puede realizar mediante distintas técnicas de muestreo que se detallan a continuación. Es fundamental conocer que el tipo de muestreo más importante es el probabilístico, en el que todos los elementos de la población tienen la misma probabilidad de ser incluidos en la muestra.

- **Muestreos probabilísticos.** En ellos se conoce la probabilidad de que un individuo sea elegido para la muestra. Son interesantes para usar estadística matemática con ellos. La elección de la muestra se hace aleatoriamente. La estadística inferencial exige este tipo de muestreo.
- **Muestreos no probabilísticos.** En ellos no se conoce la probabilidad. Son muestreos que muy posiblemente esconden sesgos. En principio, no se pueden extrapolar los resultados a la población.

Los muestreos con la menor posibilidad de sesgo (probabilísticos) son los siguientes: aleatorio simple, aleatorio estratificado, sistemático, y por grupos o conglomerados.

- **Muestreo aleatorio simple.** Es aquel en el que cada individuo tiene las mismas posibilidades de ser elegido para formar parte de la muestra.
- **Muestreo aleatorio estratificado.** En este tipo de muestreo la población se divide en estratos. La muestra se obtiene seleccionando de cada estrato un número proporcional a la población que representa. Con este tipo de muestreo se asegura que la muestra tenga la misma proporción de una(s) variable(s) que la población de la que procede y que podría tratarse de una variable-pronóstico diferente en los grupos de estudio **MIR 16-17, 130**.
- **Muestreo sistemático.** El proceso de selección empleado se basa en alguna regla sistemática simple como, por ejemplo, elegir uno de cada n individuos.
- **Muestreo por conglomerados.** En este tipo, la selección se realiza fundamentalmente con el objetivo de reducir costes, utilizando como base del muestreo al grupo de sujetos.

Por último, **individuo** será cada uno de los componentes de la población y de la muestra. Al número de individuos que forman la muestra se le llama **tamaño muestral** (n).

8.3. Variables

Una **variable** es una característica observable que se desea estudiar en una muestra de individuos, pudiendo tomar diferentes valores. Los diversos tipos de variables están recogidos en la **Tabla 8.1** **MIR 14-15, 184**.

VARIABLE	SUBTIPO	EJEMPLOS
Cualitativa	Nominal dicotómica	Sexo
	Nominal no dicotómica	Raza
	Ordinal	Nivel socioeconómico
Cuantitativa	Discreta	N.º episodios de asma/semana
	Continua	Uricemia

Tabla 8.1. Tipos de variables.

8.4. Representación gráfica

La representación gráfica de los datos obtenidos en cualquier estudio facilita un análisis visual. Según la naturaleza de las variables estudiadas, se utilizan diferentes tipos de representación.

Variables cualitativas

Se emplean:

- **Diagrama de barras (Figura 8.1).**
- **Diagrama sectorial (pastel) (Figura 8.2).**

En ambos casos, se debe cumplir el principio de proporcionalidad de las áreas a las frecuencias absolutas. En los diagramas de sectores, el ángulo central es proporcional a la frecuencia absoluta correspondiente, por lo que también lo es su área.

Los diagramas de barras tienen una base constante y una altura proporcional a la frecuencia absoluta correspondiente (también su área lo es a la frecuencia absoluta).

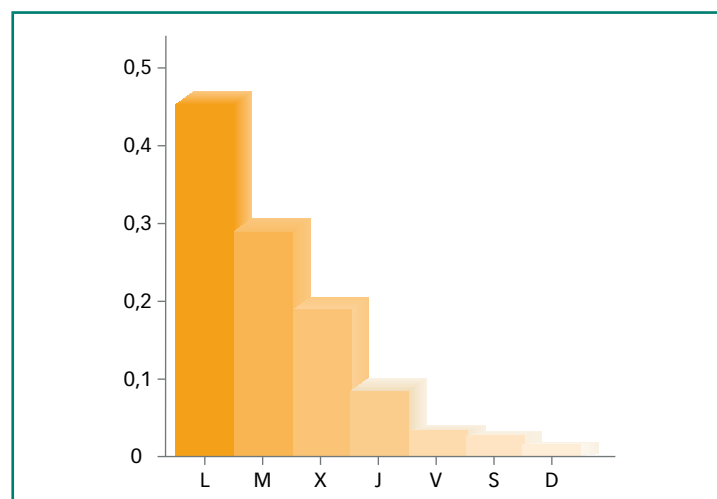


Figura 8.1. Diagrama de barras.

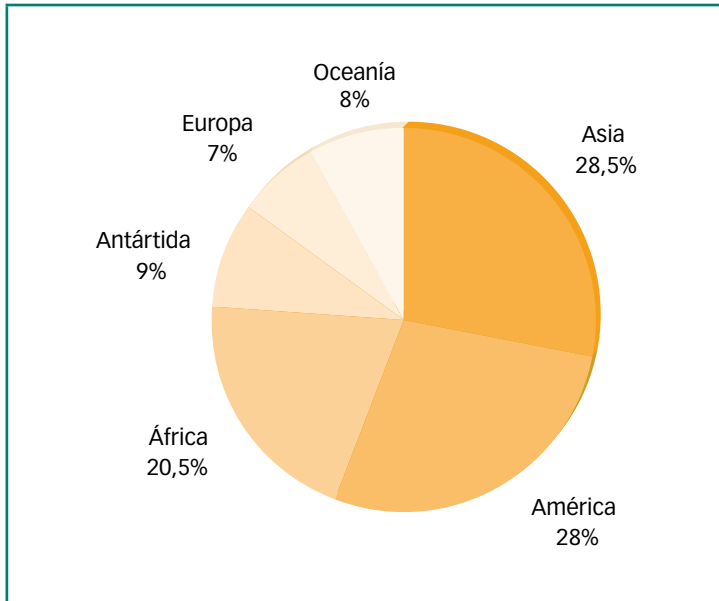


Figura 8.2. Diagrama sectorial.

■ Variables cuantitativas discretas

- **Diagrama de barras.** En un diagrama, sobre el valor que puede tomar la variable, se levanta una barra cuya altura mide exactamente la frecuencia absoluta del valor. También es posible representar la frecuencia relativa y/o los porcentajes.

■ Variables cuantitativas continuas

- **Histograma.** Es un gráfico que está formado por rectángulos adyacentes que tienen por base cada uno de los intervalos, y por altura las frecuencias absolutas. La superficie de cada rectángulo es proporcional a la frecuencia de cada una de las clases, y el área total lo será al número de individuos en la muestra (Figura 8.3).
- **Polígono de frecuencias.** Es una línea quebrada que une los puntos medios de las barras superiores de los rectángulos del histograma (Figura 8.3).

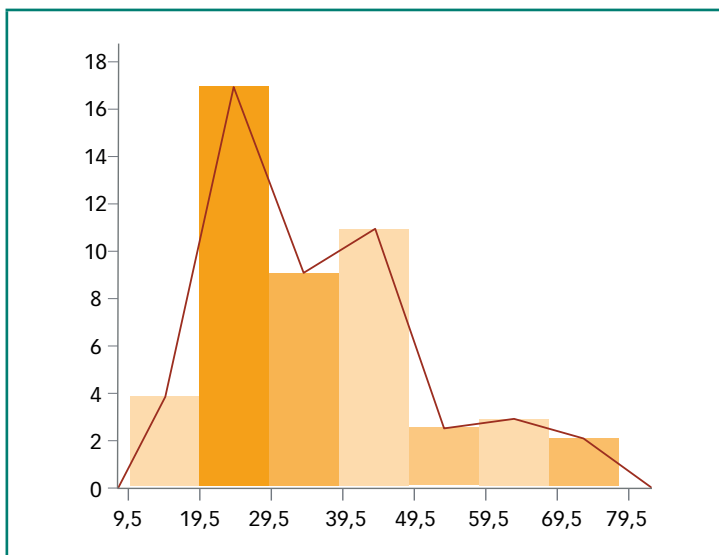


Figura 8.3. Histograma y polígono de frecuencias.

8.5. Síntesis de los datos

■ Medidas de localización o tendencia central

Las **medidas de localización** indican alrededor de qué valores se agrupan los datos observados. Se distinguen las siguientes:

- **Media aritmética.** Es la medida de centralización más común. Se calcula sumando los valores numéricos de todas las observaciones y dividiendo el total por el número de observaciones.

$$\bar{X} = \frac{\sum x_i}{n}$$

La media aritmética verifica la propiedad de equilibrar las desviaciones positivas y negativas de los datos respecto a su valor, es decir, $\sum (x_i - \bar{x}) = 0$. Actúa, por tanto, como centro geométrico o centro de "gravitación" para el conjunto de puntos.

- **Mediana.** Es el valor numérico que divide al conjunto de datos ordenados en dos partes iguales, es decir, el 50% de los datos será menor que ella y el resto, mayor. En una distribución simétrica, la mediana coincide con la media aritmética, pero no así en una asimétrica.
- **Moda.** Es el valor más corriente o el valor de la variable que se presenta con mayor frecuencia. Pueden existir distribuciones con más de una moda.
- **Parámetros de posición.** Cuartiles, deciles, percentiles. Son valores que dividen el conjunto de las observaciones en 4, 10 o 100 partes iguales, respectivamente. Ejemplo:

$$Q_2 = D_5 = Pc_{50} = \text{mediana}$$

El 50% de las observaciones será inferior al segundo cuartil, quinto decil o percentil 50.

- **Media geométrica.** Muy utilizada en microbiología y serología, cuyos datos tienen una marcada asimetría positiva (hacia la derecha). Por ejemplo, títulos de anticuerpos.

Comparación de medidas de centralización

Aunque desde un punto de vista puramente descriptivo, estas medidas proporcionan información complementaria, sus propiedades son muy distintas: la media utiliza todos los datos y es, por tanto, preferible si los datos son homogéneos; tiene el inconveniente de que es muy sensible a observaciones atípicas, y un error de datos o un valor anormal puede modificarla totalmente. Por el contrario, la mediana emplea menos información que la media, ya que solo tiene en cuenta el orden de los datos y no su magnitud, pero, en contrapartida, no se ve alterada si una observación (o, en general, una pequeña parte de observaciones) es extrema o contiene errores grandes de medida o de transcripción.

En consecuencia, es siempre recomendable calcular la media y la mediana, puesto que ambas medidas diferirán mucho cuando la distribución sea muy asimétrica (Figura 8.4), lo que sugiere heterogeneidad en los datos.

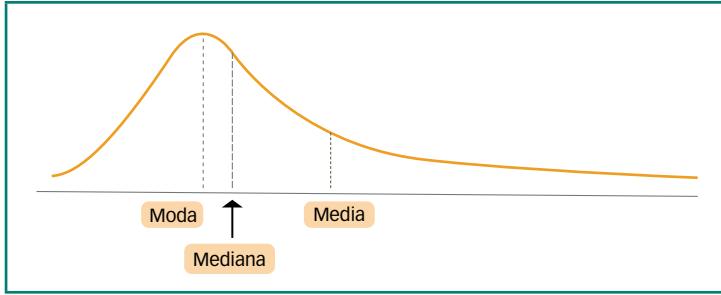


Figura 8.4. Distribución asimétrica.

■ Medidas de dispersión o variabilidad

Junto a las medidas de tendencia central (Tabla 8.2), las medidas de dispersión o variabilidad completan la información sobre la distribución de la variable (indican si los valores de la misma están muy dispersos o se concentran alrededor de la medida de centralización).

- **Rango o recorrido.** Diferencia entre el valor máximo y el mínimo observado en una serie.

$$R = \text{Máx.} - \text{Mín.}$$

- **Desviación media.** Es la media de las desviaciones respecto a la media aritmética.

$$D_m = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

- **Varianza.** Se define como la media del cuadrado de las desviaciones de los elementos respecto a la media aritmética.

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- **Desviación típica o estándar.** Se define como la raíz cuadrada positiva de la varianza. Es, junto con esta, la medida de dispersión más utilizada **MIR 13-14, 190**.

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

La desviación típica es una medida complementaria de la media aritmética; mientras que esta da una idea de la magnitud general de la distribución, la desviación estándar muestra cómo se distribuyen los valores alrededor de la media.

- **Rango intercuartílico.** Es la diferencia entre el percentil 75 y el 25. Es, junto con el rango, la medida de dispersión que se utiliza para los datos asimétricos.
- **Coefficiente de variación (CV).** Es una medida de dispersión adimensional. Se define como el porcentaje que representa la desviación estándar sobre la media. Es el método de elección para comparar la variabilidad o dispersión relativa de variables que estén expresadas en las mismas o en diferentes unidades.

$$CV = \frac{S}{\bar{X}} \times 100$$

	MEDIDA DE CENTRALIZACIÓN	MEDIDA DE DISPERSIÓN
Distribución homogénea	Media	Desviación típica o estándar
Distribución asimétrica	Mediana	Rango intercuartílico

Tabla 8.2. Formas de medida.

■ Propiedades de la media y la varianza

Para las propiedades de la media y la varianza hay que tener en cuenta que:

- Si a todos los valores de una distribución se les suma una constante, su media queda aumentada en ese valor, mientras que su varianza no se modifica.
- Si a todos los valores de una distribución se les multiplica por una constante, su media y su desviación típica quedan multiplicadas por la constante, mientras que su varianza queda multiplicada por el cuadrado de esa constante.

8.6. Asimetría y curtosis

■ Asimetría

Para calcular la asimetría, una posibilidad es utilizar el llamado **coeficiente de Fisher** que se representará como g_1 . Según sea el valor de g_1 , se dirá que la distribución es asimétrica a derechas o positiva, a izquierdas o negativa, o simétrica, es decir (Figura 8.5):

- Si $g_1 > 0$, la distribución será asimétrica positiva o a derechas (desplazada hacia la derecha).
- Si $g_1 < 0$, la distribución será asimétrica negativa o a izquierdas (desplazada hacia la izquierda).
- Si $g_1 = 0$, la distribución será simétrica.

■ Curtosis (apuntamiento)

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución.

Por medio del coeficiente de curtosis es posible identificar si existe una gran concentración de valores (leptocúrtica), una concentración normal (mesocúrtica) o una baja concentración (platicúrtica) (Figura 8.6).

El valor g_2 representa el **coeficiente de curtosis**. Si $g_2 = 0$, entonces la distribución es mesocúrtica. Al igual que en la asimetría, es bastante difícil encontrar un coeficiente de curtosis de cero (0), por lo que se suelen aceptar los valores cercanos ($\pm 0,5$ aproximadamente).

- Si $g_2 > 0$, la distribución será leptocúrtica.
- Si $g_2 < 0$, la distribución será platicúrtica.
- Si $g_2 = 0$, la distribución será mesocúrtica.

Cuando la distribución de datos tiene un coeficiente de asimetría ($g_1 = \pm 0,5$) y un coeficiente de curtosis ($g_2 = \pm 0,5$), se denomina curva normal.

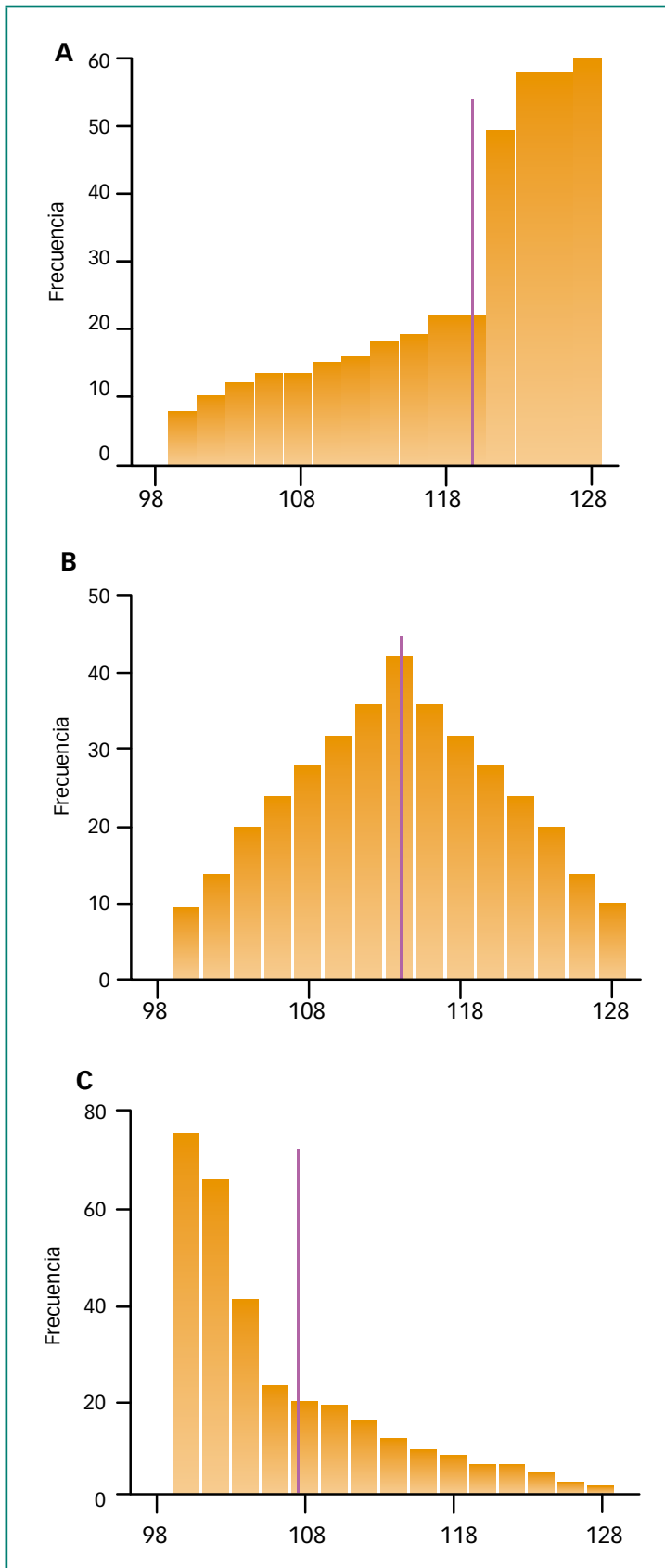


Figura 8.5. Estados de la asimetría. (A) Distribución asimétrica negativa. (B) Distribución simétrica. (C) Distribución asimétrica positiva.

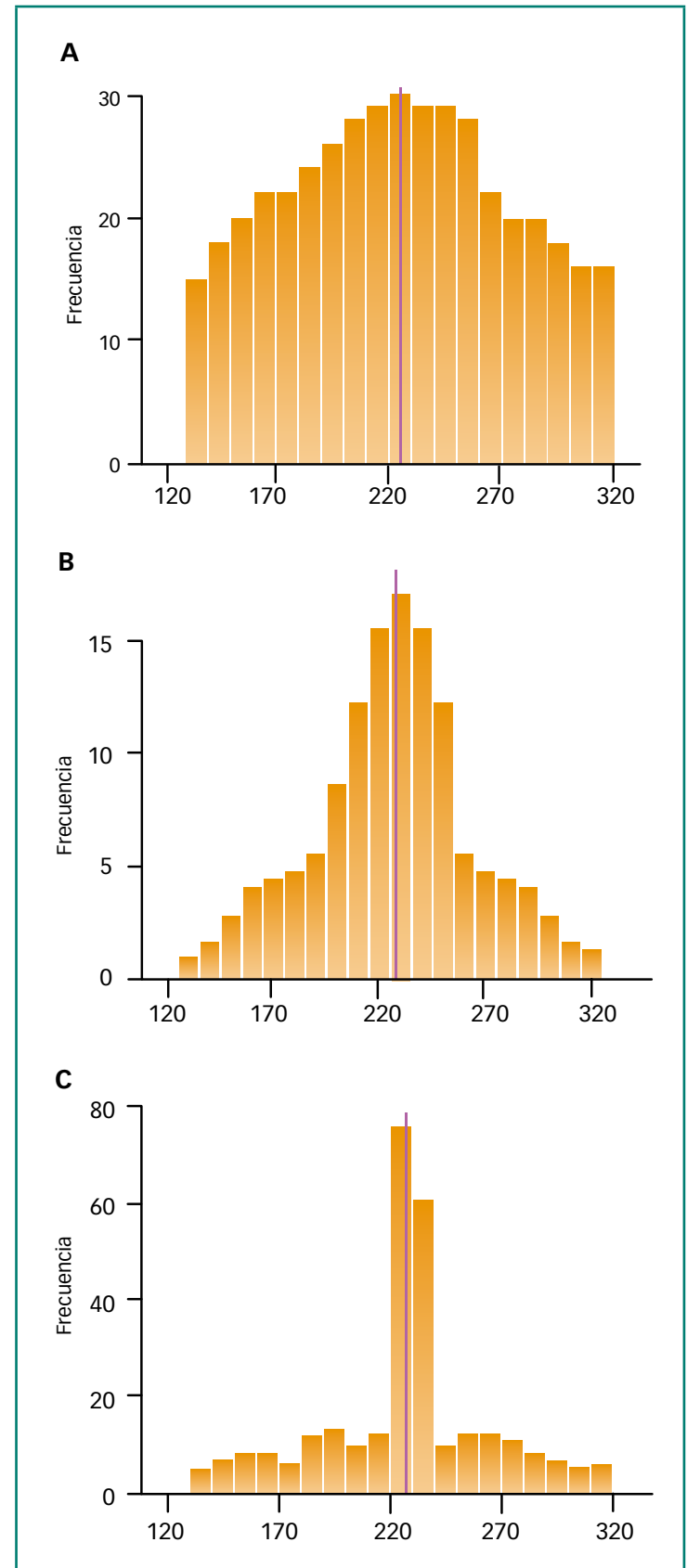


Figura 8.6. Coeficiente de curtosis. (A) Distribución platikúrtica. (B) Distribución mesokúrtica. (C) Distribución leptokúrtica.

Casos clínicos

Si se identifican las áreas sanitarias de una provincia, se eligen tres de ellas al azar, y a partir de cada una de las elegidas se selecciona un número determinado de sujetos, también al azar, se estará realizando un muestreo:

- 1) Aleatorio simple.
- 2) Aleatorio estratificado.
- 3) Sistemático.
- 4) Por conglomerados.

RC: 4

De los siguientes estadísticos de tendencia central y dispersión, indique aquel que puede tomar varios valores diferentes dentro de una misma distribución:

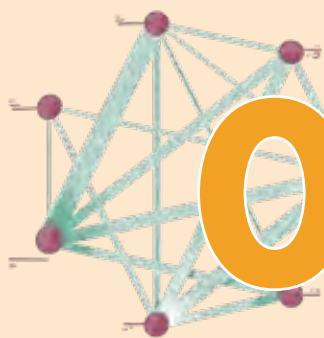
- 1) Media aritmética.
- 2) Mediana.
- 3) Moda.
- 4) Varianza.

RC: 3

Las medidas de tendencia central informan sobre los valores centrales hacia los que tiende la distribución, mientras que las medidas de dispersión dan una idea de cómo de agrupados o dispersos están los valores de la muestra respecto a los parámetros centrales. Un estudio pretende comparar la dispersión relativa del peso y talla de un grupo de ancianos. ¿Qué medida de dispersión emplearía?

- 1) Varianza.
- 2) Desviación típica.
- 3) Rango.
- 4) Coeficiente de variación.

RC: 4



Principales leyes de distribución de variables aleatorias

Orientación MIR

Uno de los temas importantes del MIR es saber interpretar intervalos de una muestra a partir de la curva de Gauss o normal. Hay que conocer perfectamente esta distribución.

Preguntas MIR

- ▷ MIR 18-19, 122; MIR 18-19, 123
- ▷ MIR 17-18, 215
- ▷ MIR 14-15, 185

Conceptos clave

- ⊙ La distribución normal o de Gauss es una curva de probabilidad continua con forma de campana.
- ⊙ Los tres parámetros de tendencia central son: media, mediana y moda, y coinciden en el mismo valor.
- ⊙ Los límites de dicha campana son $\pm\infty$ y su área es 1.
- ⊙ Siempre se cumple que en el intervalo $\text{media} \pm 1,96$ s está incluido el 95% de la muestra.
- ⊙ La estadística inferencial o analítica se encarga de, a partir de la muestra, extrapolar datos a la población.
- ⊙ El parámetro de dispersión que siempre se emplea para el cálculo de los intervalos de confianza de medias es el error estándar de la media.
- ⊙ La curva de probabilidades de los intervalos de confianza sigue otra curva de Gauss, por lo que se puede decir que, en el intervalo $\text{media} \pm 2$ eem, estará incluida la media poblacional con una probabilidad del 95%.
- ⊙ Los intervalos así obtenidos se denominan intervalos de confianza del 95%.
- ⊙ El 5% restante es la probabilidad de que el intervalo esté mal estimado. A este error se le llama α .

Una vez que se ha elegido la muestra y recogido los datos, el siguiente paso es inferir las propiedades de la población a partir de la muestra.

El instrumento conceptual que permitirá esta generalización es un modelo de la población, es decir, una representación simbólica de su comportamiento.

Existen tres leyes teóricas que tienen la particularidad de ser seguidas por una inmensa mayoría de los fenómenos biológicos habituales; estas leyes son:

- Distribución binomial (distribuciones discretas).
- Distribución de Poisson o de los sucesos raros (distribuciones discretas).
- Distribución normal o de Gauss (distribuciones continuas).

9.1. Distribución binomial

Es una ley teórica de aplicación siempre que se conozca, de entrada, la probabilidad de aparición de un fenómeno biológico (p). El resultado es dicotómico y puede ser evaluado como "éxito o con fracaso". La variable de interés es el número de éxitos alcanzados en n pruebas.

La distribución binomial constituye siempre un diagrama de barras discontinuo que se aplica a variables discretas.

9.2. Distribución de Poisson

En general, se empleará la distribución de Poisson como aproximación de experimentos binomiales donde el número de pruebas es muy alto, pero la probabilidad de éxito muy baja. Es, por consiguiente, la ley adecuada para los "sucesos raros", cuando $p < 0,1$ y $n \times p < 5$.

9.3. Distribución normal o de Gauss

Este tipo de distribución es seguida por una inmensa cantidad de variables biológicas cuyas medidas se agrupan alrededor de un valor central, y que presentan una frecuencia cada vez menor a medida que se alejan de dicho valor.

■ Características

Las características de la distribución normal (Figura 9.1) o de Gauss son las siguientes:

- Corresponde a variables cuantitativas continuas.
- Se caracteriza por dos medidas: media y desviación típica.
- Es unimodal.
- Es simétrica alrededor de la media. Por tanto, media, mediana y moda coinciden.
- Tiene forma acampanada, sin un pico excesivo.
- Va desde $-\infty$ a $+\infty$ (asintótica al eje de abscisas).
- El área bajo la curva tiene un valor igual a 1.

El 50% de las observaciones se encuentran por debajo de la media aritmética y el 50% por encima. El 68% están dentro del intervalo $\bar{x} \pm s$; el 95% dentro del intervalo $\bar{x} \pm 1,96 s$, y el 99% dentro del intervalo $\bar{x} \pm 2,57 s$.

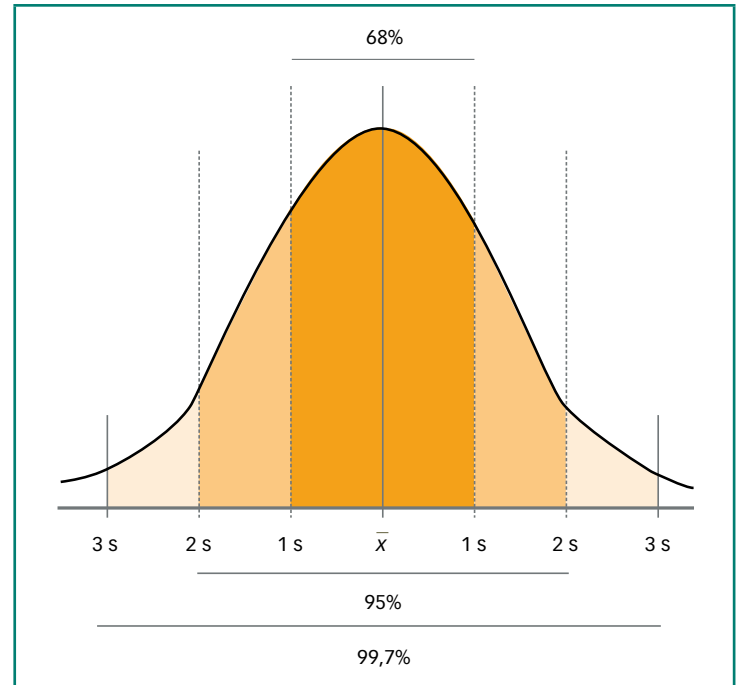


Figura 9.1. Distribución normal o de Gauss.

9.4. Estimación de medias

La tarea fundamental de la estadística inferencial es hacer inferencias acerca de la población a partir de una muestra extraída de la misma. En este caso, se quiere conocer la media poblacional a partir de la media muestral.

Para la estimación de medias se pueden aplicar dos tipos de estimación: la estimación puntual y la estimación por intervalos.

- **Estimación puntual.** La media muestral es un buen estimador de la media poblacional (μ).
- **Estimación por intervalos.** El valor numérico de los estimadores puntuales determinados en muestras diferentes puede variar. Es decir, si se repiten los cálculos con otra muestra distinta de la misma población, el resultado de la media muestral podría ser diferente. Por tanto, sería mejor que, además, se aportase un intervalo que presumiblemente incluya también el parámetro de la población. Por ello, es preferible la estimación por intervalos, ya que entonces se indican límites de valores dentro de los cuales el parámetro poblacional tiene la probabilidad de estar.

Al intervalo alrededor del estadístico muestral se le denomina intervalo de confianza, y a sus límites, límites de confianza. El cálculo de los límites de confianza (IC), comprende el empleo del error estándar de la media y los principios de la distribución normal ▶ MIR 18-19, 123.

El nivel de confianza $1 - \alpha$ indica la probabilidad de que la media poblacional, μ , pertenezca a ese intervalo; es frecuente expresarlo en porcentajes. Son equivalentes las expresiones "nivel de confianza $1 - \alpha$ " y "significación α "; así, suele hablarse de obtener un intervalo con un nivel de confianza del 95% o para una significación $\alpha = 0,05$ (este valor de α indica la probabilidad de que hay de errar en la estimación. A veces, esa probabilidad se indica con la letra p).

La amplitud de un IC depende del nivel de confianza y del tamaño muestral, siendo el IC más pequeño con un nivel de confianza menor y con un tamaño muestral mayor ▶ MIR 17-18, 215.



■ Error estándar de la media

De una población pueden extraerse infinitas muestras, cada una de ellas con su media. Este conjunto de medias se distribuye según una curva normal cuya media, la media de las medias, es la media poblacional (μ) y cuya desviación típica se conoce como **error estándar de la media**; por tanto, es la dispersión de las medias muestrales con respecto a la media poblacional.

Por la teoría central del límite, aunque una distribución muestral no se ajusta a la normal, asumimos que la estimación de medias en la población sí se ajustará a dicha distribución.

Se calcula con la siguiente fórmula (Figura 9.2) ► MIR 18-19, 122; MIR 14-15, 185 :

$$EE_m = \frac{s}{\sqrt{n}}$$

Si se construye un IC del 95%, se concluye que en el 95% de las ocasiones la media poblacional estará entre:

$$\bar{x} \pm 1,96 \times \frac{s}{\sqrt{n}}$$

Mientras que en un 5% de las ocasiones estará equivocado. Si se establece el nivel de confianza en el 99%, la media poblacional se situará en un 99% de las ocasiones entre:

$$\bar{x} \pm 2,57 \times \frac{s}{\sqrt{n}}$$

En el caso de $n < 30$, los límites de confianza se calcularían siguiendo la fórmula:

$$\bar{x} \pm t \times \frac{s}{\sqrt{n}}$$

Donde el valor t corresponde a los valores tabulados de la distribución teórica t de Student.

Recuerda

- No tiene nada que ver la desviación típica de los valores de la variable de la muestra o en la población con el error estándar de las medias muestrales alrededor de la media poblacional.

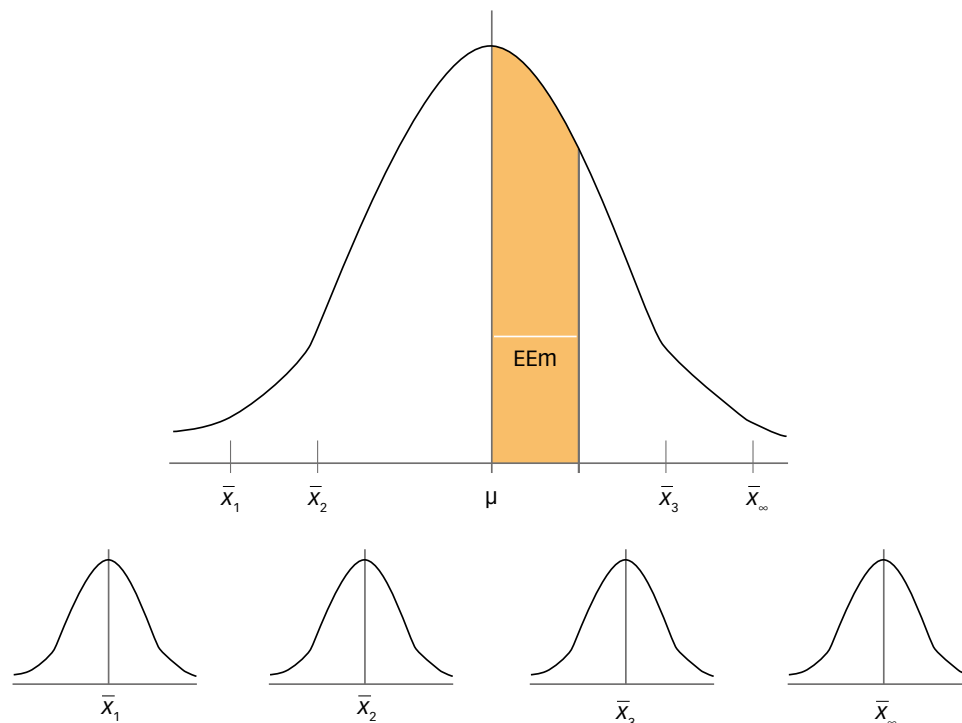


Figura 9.2. Cálculo del error estándar de la media.

Casos clínicos

Las edades de una muestra de 300 pacientes siguen una distribución normal con media de 50 años y desviación típica 10 años. Esto implica que:

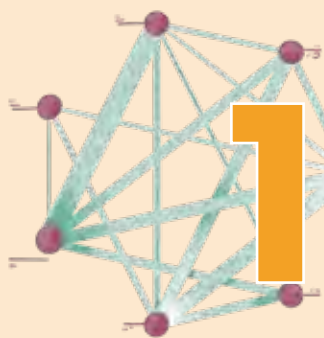
- 1) Aproximadamente el 95% de los pacientes tienen edades entre 40 y 60 años.
- 2) Existe un 95% de probabilidades de que la verdadera edad media se sitúe entre 40 y 60 años.
- 3) Existe un 95% de probabilidades de que la verdadera edad media se sitúe entre 30 y 70 años.
- 4) Aproximadamente el 95% de los pacientes tienen edades entre 30 y 70 años.

RC: 4

En la estimación estadística de una media poblacional para una determinada variable:

- 1) Los intervalos de confianza que se calculan para la media poblacional son independientes de la variación en el muestreo.
- 2) El error estándar de la media es independiente del tamaño de la muestra.
- 3) La distribución de medias muestrales tiende a ser una distribución normal siempre que las muestras sean todas del mismo tamaño, procedan aleatoriamente de la misma población y se mida una variable cuantitativa continua.
- 4) Un intervalo de confianza del 99% supone un nivel de significación de $p \leq 0,05$.

RC: 3



El contraste de hipótesis

Orientación MIR

Lo que hay que tener claro de este tema es el significado de los errores α y β , y del poder de un test estadístico. También es muy frecuente que pregunten sobre cómo se modifican los parámetros anteriores en función del tamaño de la muestra.

Preguntas MIR

- ▷ MIR 18-19, 115
- ▷ MIR 17-18, 214
- ▷ MIR 16-17, 122; MIR 16-17, 123; MIR 16-17, 124
- ▷ MIR 15-16, 194
- ▷ MIR 14-15, 186
- ▷ MIR 13-14, 33; MIR 13-14, 34; MIR 13-14, 192; MIR 13-14, 204

Conceptos clave

- ◉ El contraste de hipótesis se basa en intentar rechazar H_0 (no hay diferencias) para así aceptar H_1 (sí existen diferencias).
- ◉ Cualquier resultado observado en un estudio puede atribuirse al azar. El valor de p , o nivel de significación, se refiere a la probabilidad de que el azar explique un resultado obtenido.
- ◉ Si $p < 0,05$, se admite poca probabilidad de azar, y entonces se habla de que hay diferencias estadísticamente significativas. Si $p > 0,05$, la probabilidad de azar es demasiado grande como para que se afirme que hay diferencias significativas en los resultados obtenidos.
- ◉ La probabilidad de que se afirme que hay diferencias, cuando en realidad no las hay, equivale a la probabilidad p y se llama error α o tipo I.
- ◉ Por el contrario, si verdaderamente existen diferencias y no se detectan, se estará incurriendo de nuevo en un error, que en este caso es el β o tipo II.
- ◉ Ambos errores α y β son dos formas diferentes de errores aleatorios, por lo que se podrán reducir si se aumenta el tamaño de la muestra.
- ◉ Si verdaderamente hay diferencias reales en los resultados obtenidos y estas diferencias son detectadas por el test, entonces se está en la situación de acierto, que se conoce como poder o potencia del test.
- ◉ Los test estadísticos que se van a emplear dependerán del tipo de variable que se haya utilizado:
 - Cuali + cuali: χ^2 al cuadrado.
 - Cuanti + cuali dicotómica: t de Student (equivalente no paramétrico: Wilcoxon, U de Mann-Whitney).
 - Cuanti + cuali no dicotómica: ANOVA (equivalente no paramétrico: Kruskal-Wallis, Friedman).
 - Cuanti + cuanti: correlación de Pearson (equivalente no paramétrico: Spearman).
- ◉ Los test paramétricos son preferibles, pero únicamente pueden realizarse cuando la distribución de estudio sea la normal.
- ◉ Los test de correlación y de regresión sirven para establecer asociación entre dos variables cuantitativas.
- ◉ El análisis del test de correlación se basa en la interpretación del coeficiente r . El valor de r está comprendido entre -1 y $+1$. Un valor positivo significa que si la variable n aumenta, la variable m lo hará también. El valor absoluto de la r establece lo fuerte o débil de la correlación, considerándose fuerte cuando esta es mayor de $0,7$.
- ◉ Una r de 1 o de -1 implica una correlación cuya nube de dispersión se aproxima a la línea recta.
- ◉ Cuando la correlación es fuerte, podrá entonces hacerse el test de regresión lineal, cuya fórmula es $y = \alpha + \beta x$, donde β es la pendiente de la recta y α el punto de corte de esta en el eje de ordenadas.

10.1. Hipótesis nula, hipótesis alternativa y grado de significación estadística

Para la comprensión de este concepto, se partirá de un caso práctico. Se compara un tratamiento nuevo (N) contra la hipertensión arterial con uno tradicional (T) en dos series de individuos. Tras cierto tiempo, N controla al 70% de los individuos y T al 30%. El problema consiste en saber si este hecho es debido a una mayor eficacia real de N o lo es simplemente por diferencias que el azar puede establecer (variaciones aleatorias del muestreo), sin que existan diferencias reales en el control de hipertensos entre ambos tratamientos.

Para abordar este problema, se consideran dos hipótesis:

- **Hipótesis nula (H_0):** no existen diferencias entre los dos tratamientos ($N = T$).
- **Hipótesis alternativa (H_1):** sí existen diferencias entre los dos tratamientos ($N \neq T$).

Estas dos hipótesis son mutuamente excluyentes, de forma que solo hay dos decisiones posibles:

- Rechazar $H_0 \rightarrow$ aceptar H_1 .
- No rechazar $H_0 \rightarrow$ no poder aceptar H_1 .

Previamente al ensayo de una hipótesis, se fija la probabilidad máxima de que los resultados diferentes observados entre los dos grupos puedan ser debidos simplemente al azar (H_0 cierta), que suele ser por convenio del 5%. A continuación, se calcula cuál es la probabilidad de que las diferencias que se han observado puedan ser explicadas por azar. Esta probabilidad es el valor de la p o "grado de significación estadística". Así, cuanto menor sea p , esto es, cuanto más baja sea la probabilidad de que el azar sea el responsable de las diferencias, mayor será la evidencia contra H_0 y a favor de H_1 .

De tal forma, si:

- $p < 0,05 \rightarrow$ diferencias reales. Existe poca probabilidad de que se deban al azar. Se acepta H_1 . El resultado es estadísticamente significativo. Se rechaza H_0 ► MIR 13-14, 204.
- $p > 0,05 \rightarrow$ no existe suficiente evidencia como para decir que ambos tratamientos son distintos. Las diferencias pueden deberse al azar, con una probabilidad mayor al nivel de exigencia. No se rechaza H_0 .

El valor p es, por tanto, la probabilidad de observar los resultados de un estudio (u otros más alejados de H_0) si la H_0 fuese cierta.

Es muy habitual caer en el error de considerar que la p es la probabilidad de que H_0 sea cierta. Se parte de la hipótesis de que H_0 es cierta para poder calcular después el valor de la p (p. ej., si se tiene un estudio en el que la p es 0,03, no se podrá decir nunca que hay una probabilidad del 3% de que los dos grupos sean iguales).

La característica que más influye en el cálculo de los valores de la p es el tamaño de la muestra. Con muestras de pequeño tamaño ($n < 30$) es muy probable que aparezcan valores de p no significativos. Sin embargo, con tamaños muestrales muy grandes es frecuente encontrar diferencias por pequeñas que estas sean.

10.2. Errores α y β

- **Error tipo I.** Habitualmente del 5%. Se rechaza H_0 siendo cierta (se acepta que hay diferencias y, de hecho, no las hay). Es como un falso positivo: dar como significativo algo que no lo es. Se denomina α a la probabilidad de cometer el error tipo I. El valor de p coincide con la probabilidad de cometer el error tipo I. Cuando entre los distintos grupos de estudio se compara más de una variable de resultado, aumenta la posibilidad de resultados falsamente positivos. Para evitarlo, se incrementa la exigencia del nivel de significación como sigue:

$$p = \frac{0,05}{N.º \text{ de comparaciones}}$$

- **Error tipo II.** Habitualmente del 20%. No se rechaza H_0 y esta es falsa (no se aceptan las diferencias y sí las hay). Sería como un falso negativo: se da como no significativo algo que lo es. Se denomina β a la probabilidad de cometer un error tipo II ► MIR 18-19, 115.
- **Poder o potencia del test.** Habitualmente del 80%. Lo complementario del error β es la "potencia o poder estadístico de un test" ($1 - \beta$), que se define como la capacidad que tiene un test de detectar una diferencia cuando esta existe en realidad; es decir, corresponde a la probabilidad que se tiene de demostrar la hipótesis alternativa, siendo cierta (Tabla 10.1).

		REALIDAD	
		Existe diferencia H_0 falsa	No existe diferencia H_0 cierta
Resultados del test	Hay diferencias significativas Rechazo H_0	$1 - \beta$ Poder estadístico o potencia del test	Error tipo I o error α
	No hay diferencias significativas No rechazo H_0	Error tipo II o error β	$1 - \alpha$

Tabla 10.1. Contraste de hipótesis.

10.3. Pruebas de significación estadística

Las pruebas de significación estadística se resumen en la Tabla 10.2.

Las pruebas de significación estadística o test de contraste de hipótesis pretenden demostrar o no la asociación estadística entre dos variables observadas en una muestra, generando un valor de p . Obviamente, dependiendo del tipo de variables que se tenga que analizar, se debe utilizar un test de contraste de hipótesis u otro.

- **Datos apareados.** En una única muestra se estudia si existe una diferencia significativa en la variable-resultado antes y después de algún acontecimiento relevante.
 - Variable-resultado cualitativa de dos o más categorías: test de McNemar.
 - Variable-resultado cuantitativa: t de Student apareada.
 - Variable-resultado cuantitativa (varias medidas repetidas): ANOVA para medidas repetidas.



VARIABLE PRONÓSTICA O INDEPENDIENTE	VARIABLE FINAL O DEPENDIENTE	TEST	OBSERVACIONES
Cualitativa	Cualitativa	Ji al cuadrado ► MIR 16-17, 122	Muestras grandes
		Test exacto de Fisher	Muestras pequeñas
		Test de McNemar	AP
		Regresión logística	Multivariante
Cualitativa	Cuantitativa	t de Student	Cuantitativa en 2 grupos
		ANOVA ► MIR 16-17, 124	Cuantitativa en > 2 grupos
		Mann-Whitney ► MIR 16-17, 123	Cuantitativa 2 grupos. NP
		Wilcoxon	Cuantitativa 2 grupos. NP. AP
		Kruskal-Wallis	Cuantitativa > 2 grupos. NP
		Friedman	Cuantitativa > 2 grupos. NP. AP
Cuantitativa	Cuantitativa	Regresión lineal	
		Correlación de Pearson ► MIR 14-15, 186	
		Correlación de Spearman	NP
		Regresión múltiple	Multivariante
Cualitativa	Tiempo hasta cualitativa (supervivencia)	Kaplan-Meier	Método descriptivo
		Log-Rank	Compara curvas de supervivencia
		Breslow	Compara curvas de supervivencia
		Regresión de Cox	Multivariante

AP: datos apareados; NP: no paramétrico.

Tabla 10.2. Pruebas de contraste de hipótesis más habituales en medicina.

- **Test no paramétricos.** Son pruebas que no dependen de la distribución de la variable y no precisan condición previa de normalidad (**Figura 10.1**). Estas pruebas no manejan los valores cuantitativos que toma la variable en cuestión, sino únicamente sus rangos. Son pruebas de menor precisión que las paramétricas, por lo que solamente demostrarán como significativas diferencias mayores. Por tanto, generalmente se prefiere utilizar pruebas paramétricas (t de Student, análisis de la varianza, etc.) y el uso de los test no paramétricos se limita a estudios con tamaños muestrales menores de 30, en los que las poblaciones no puedan ser consideradas como normales, o bien cuando la variable represente únicamente una medida aproximada del carácter, sin una valoración exacta. Si las muestras son mayores de 30, no existe inconveniente en utilizar pruebas paramétricas.
- **Análisis multivariante.** Pretende establecer la relación entre dos variables teniendo en cuenta la influencia de otras. Tiene sentido utilizar este análisis en medicina, al encontrarse muchas características relacionadas unas con otras. Por ejemplo, si se quiere valorar la existencia de asociación entre la hipertensión arterial y el riesgo de ACV, habrá que controlar (o analizar) el efecto que pueda tener la presencia o no de dia-

betes, dislipidemia, el sexo o la edad. Se desarrollará más ampliamente este concepto más adelante.

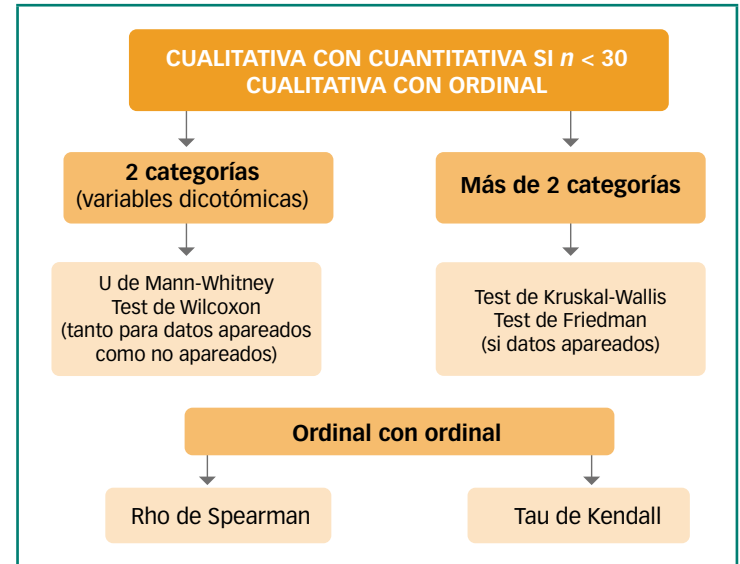


Figura 10.1. Test no paramétricos.

- **Análisis de supervivencia.** En ocasiones, para saber si, por ejemplo, un fármaco es mejor que otro, se necesita conocer la idea dinámica de cómo fallecen los pacientes incluidos en uno u otro grupo de tratamiento a lo largo del tiempo. Es decir, no solo importa si fallecen o no, si no el tiempo en el que se produce el evento.

Esta idea es extrapolable a otras situaciones como, por ejemplo, tiempo hasta que se produce la curación, tiempo hasta que desaparecen los síntomas... Uno de los análisis estadísticos empleados con más frecuencia es el de Kaplan-Meier (**Figura 10.2**). La característica distintiva del análisis con este método es que la proporción acumulada que sobrevive se calcula para el tiempo de supervivencia individual de cada paciente. Para saber si existen diferencias en el tiempo en el que se produce el evento para los dos grupos de tratamiento, hay que usar test específicos para tal fin, como el de Breslow o el de Log-Rank (**Tabla 10.3**) ► MIR 17-18, 214; MIR 13-14, 33; MIR 13-14, 34 .

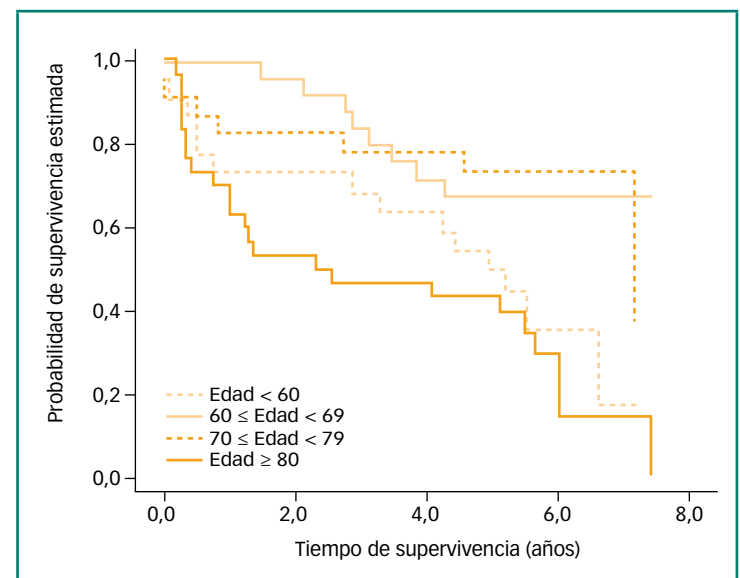


Figura 10.2. Ejemplo de test de Kaplan-Meier (KM) que evalúa la probabilidad de supervivencia según la edad de los pacientes.

La **Figura 10.3** resume los test estadísticos que más se emplean y que se han comentado anteriormente.

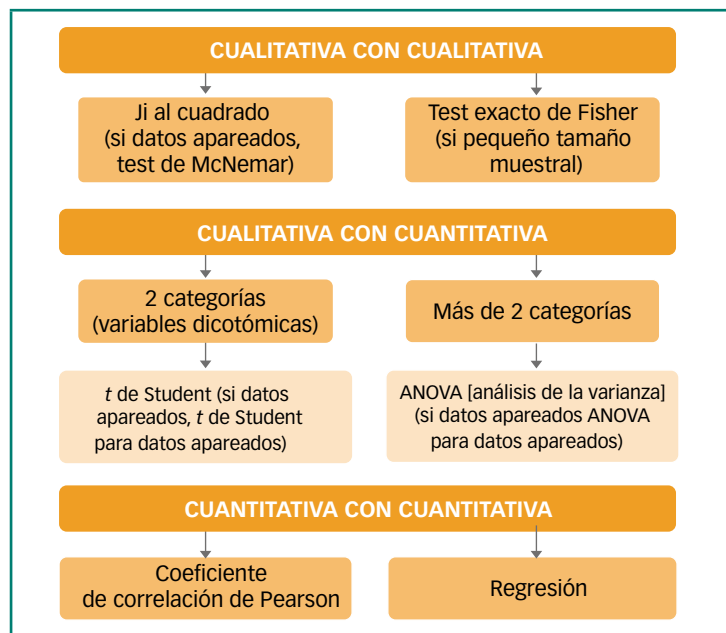


Figura 10.3. Test estadísticos más empleados.

10.4. Pruebas a una cola y a dos colas

Lo habitual de la práctica clínica es que cuando se comparan, por ejemplo, dos tratamientos, pueda ocurrir que A es mejor que B, pero también que B es mejor que A (hipótesis alternativa bilateral). En determinadas circunstancias, sin embargo, casi con toda probabilidad solo cabe la opción A mejor que B (p. ej., al comparar un antihipertensivo con placebo, hipótesis alternativa unilateral).

Los test estadísticos a dos colas (en el casos de H_1 bilateral) dan siempre valores de p mayores (en consecuencia menos significativos) que los de a una cola (en caso de H_1 unilateral). Por tanto, cuando una prueba bilateral es significativa, también lo será la unilateral. Hacer un test a una o dos colas depende fundamentalmente de los conocimientos previos (se comparan dos tratamientos o se comparan con placebo); si se tienen dudas de qué test sería conveniente emplear, se deberían usar pruebas a dos colas.

10.5. Análisis multivariante

Es un conjunto de test estadísticos que se aplica cuando se intenta establecer la asociación estadística entre dos variables, pero controlando el efecto de terceros factores que podrían actuar como factores de confusión o modificadores del efecto.

Por ejemplo, se desea saber si existe una relación entre la infección de herida quirúrgica y el tipo de cirugía, controlando que la profilaxis antibiótica empleada no actúe como factor de confusión. Para ello se trataría de establecer una expresión lógica como la siguiente:

- El riesgo de infección de herida quirúrgica depende del tipo de profilaxis y tipo de cirugía.

Incluso se podrían añadir otras muchas variables (de ahí el nombre de multivariante) como las siguientes.

- El riesgo de infección de herida quirúrgica depende del tipo de profilaxis y tipo de cirugía y edad del paciente y tiempo de cirugía... Se debe expresar esta relación de forma matemática, resultando algo así: Probabilidad (infección de herida) = Profilaxis + Tipo cirugía + Edad + Tiempo de cirugía

Hay que tener en cuenta que siempre que se relacione matemáticamente, en forma de una ecuación, una variable con otra u otras se está generando una **ecuación de regresión** **MIR 15-16, 194**. La variable que se coloca antes del igual será la variable dependiente, puesto que sus valores dependerán de los valores que tomen las variables que están después del signo igual, o variables independientes.

Existen muchos tipos de regresiones que vienen determinados por la naturaleza de la variable dependiente, esto es, si es cuantitativa, cualitativa, etc. (**Tabla 10.3**).

Al final, lo que se consigue al formular una ecuación de regresión con los datos de un estudio es generar un modelo matemático que pretende explicar el comportamiento de las variables incluidas.

Al terminar el análisis, el modelo de regresión proporcionará una estimación del riesgo asociado a cada una de las variables, y este riesgo (en forma de OR, RR o HR) es independiente para cada una de ellas.

TIPO DE VARIABLE DEPENDIENTE	CUALITATIVA	CUANTITATIVA	SUPERVIVENCIA
Ejemplo	IAM sí/no	Carga viral	Tiempo hasta oclusión de un <i>stent</i> /BP (datos de supervivencia)
Test estadístico	Regresión logística	Regresión lineal múltiple	Regresión de Cox
Medida de asociación	Odds ratio Riesgo relativo	–	Hazard ratio

Tabla 10.3. Tipos de análisis multivariantes.

10.6. Regresión lineal

La ecuación de regresión lineal (**Figura 10.4**) más sencilla (o regresión lineal simple) tiene la forma siguiente:

$$y = \alpha + \beta x$$

Donde α y β son los llamados coeficientes de regresión, y es la variable dependiente y x la variable independiente.

El coeficiente α representa el punto en el que la línea corta el eje vertical (valor de y para $x = 0$). El coeficiente β es la pendiente de la recta (puesto que la figura que genera la ecuación es una recta) que muestra la cantidad que varía y por una unidad de cambio de x .

Por ejemplo, se quiere plantear la existencia de asociación entre el índice de masa corporal (IMC) y la colesterolemia plasmática. Se generará una ecuación como esta:

$$\text{IMC} = \alpha + \beta \times \text{colesterolemia}$$

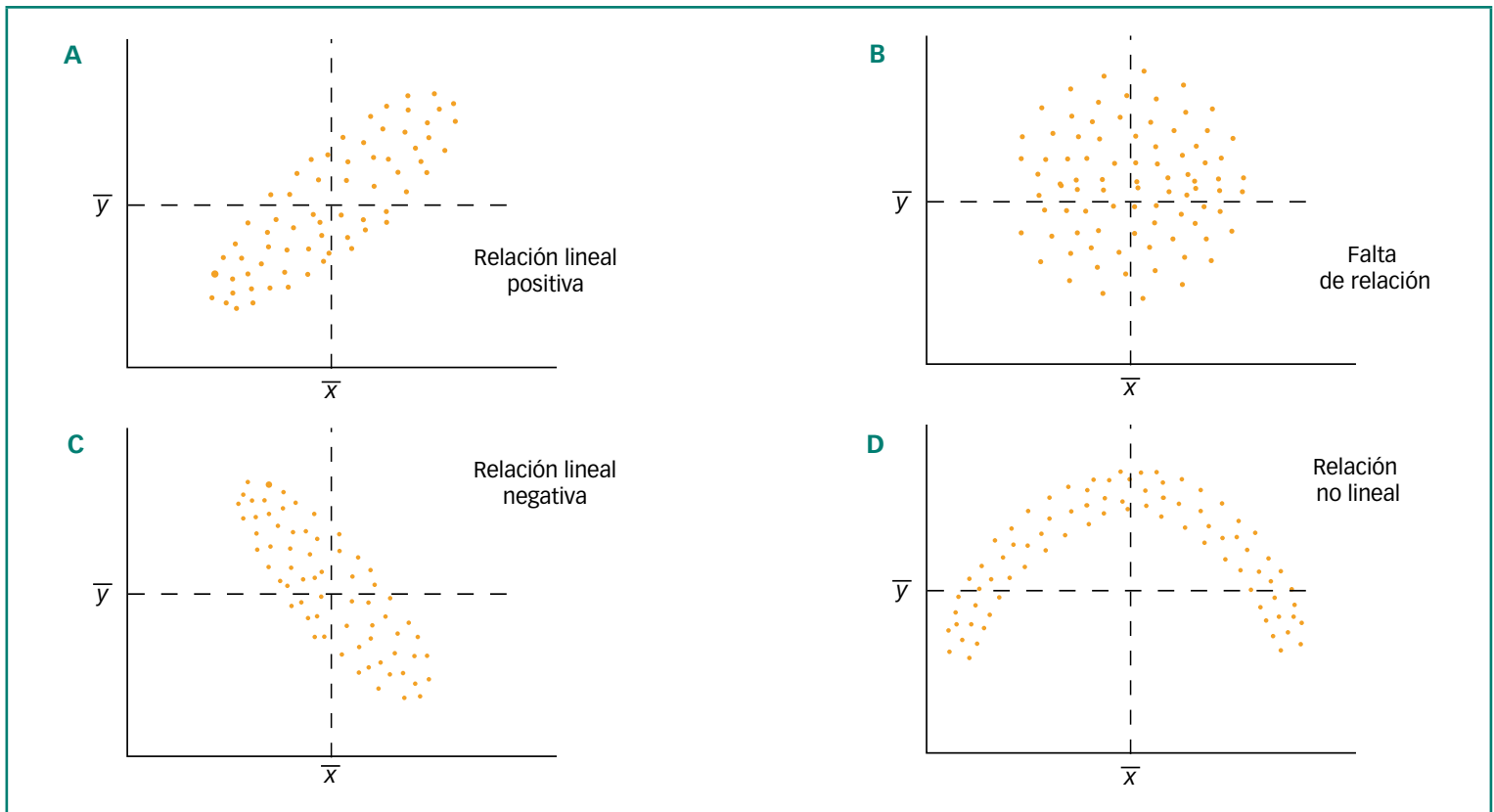


Figura 10.4. Distintos tipos de relación entre las variables.

Imagínese un resultado hipotético como este: $IMC = 0,25 + 1,86 \times \text{colesterolemia}$. El coeficiente β se interpretaría de esta forma: cada unidad que aumenta la colesterolemia (p. ej., de 200 a 201 mg/dL) el IMC aumentaría 1,86 puntos (Figura 10.5).

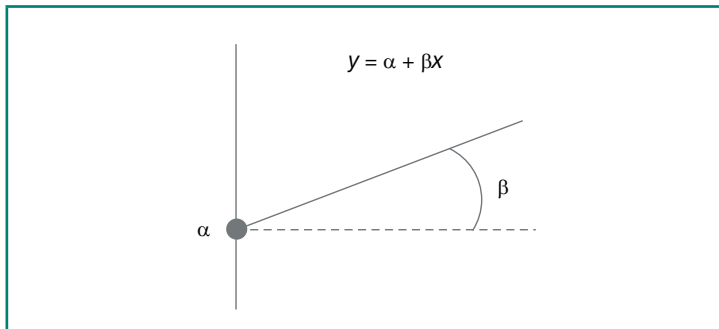


Figura 10.5. Recta de regresión lineal.

Otra forma de analizar la posible asociación entre dos variables continuas es utilizar un análisis de correlación por medio del coeficiente de correlación de Pearson ▶ MIR 14-15, 186.

Este mide la intensidad de la relación lineal entre las dos variables cuantitativas. Las características del coeficiente de correlación son las siguientes:

- Varía entre -1 y $+1 \rightarrow -1 \leq r \leq +1$.
- Para $r = \pm 1$, hay una relación perfecta entre x e y , es decir, todos los puntos (x, y) están en una línea recta.
- Un valor positivo de r indica que a medida que aumenta una variable, lo hace la otra, o que conforme disminuye una, también lo hace la otra. Un coeficiente de correlación negativo denota que a medida que disminuye una variable, aumenta la otra o viceversa.
- $r = 0$ indica que no hay correlación lineal ▶ MIR 13-14, 192.

En general, si se tiende a realizar primero un análisis de correlación, y de este resulta una asociación fuerte entre variables, se intentará formular una ecuación de regresión lineal.

10.7. Regresión logística

La ecuación de regresión logística es matemáticamente mucho más compleja, pero desde un punto de vista conceptual, es exactamente igual a la regresión lineal salvo por el tipo de variable dependiente, que es cualitativa dicotómica. El final de un análisis de regresión logística, como se ha comentado anteriormente, da lugar a una cuantificación del riesgo asociado a cada una de las variables independientes incluidas. En la Tabla 10.4 se muestra un cuadro semejante a cómo estarían presentados los resultados de un análisis de regresión logística. En esa tabla interesan los valores de Exp (B) y sus IC, puesto que representarían el valor del RR u OR y su IC, respectivamente.

Así, el riesgo de infección de herida sería de 1,03 (IC 1,021-1,041) si se hace profilaxis (es una OR o un RR, en función del diseño del estudio), de 1,024 (IC 0,733-1,034) por cada año que se aumenta la edad (no significativo) y de 4,118 (IC 1,740-9,650) por cada hora que aumenta el tiempo quirúrgico.

VARIABLE	B	E.S.	P	EXP (B)	IC 95% EXP (B)	
					INFERIOR	SUPERIOR
Profilaxis	1,641	0,291	0,003	1,03	1,021	1,041
Edad	-2,156	0,211	0,225	1,024	0,733	1,034
Tiempo de cirugía	4,109	0,537	0,000	4,118	1,740	9,650

Tabla 10.4. Modelo de presentación de los resultados de un análisis de regresión logística.

10.8. Regresión de COX

Podría considerarse un caso particular de regresión logística en la que la variable dependiente incluye el tiempo hasta que se produce un evento de interés (datos de análisis de supervivencia).

Su resultado sería el mismo que el expresado para la regresión logística, esto es, podría ser una tabla semejante a la **Tabla 10.4** anterior, en la que los valores de Exp (B) corresponderían a *hazard ratio* (HR), una medida de asociación que se puede interpretar igual que una OR.

Casos clínicos

Se realiza un estudio para comparar el efecto de dos nuevos anti-hipertensivos, A y B, sobre la tensión arterial. Para ello se diseña un ensayo clínico, correcto en su ejecución, y se realiza un contraste de hipótesis, partiendo de la hipótesis nula de igualdad. Después de aplicar el modelo estadístico necesario, se concluye con la aceptación de la hipótesis nula. Años más tarde, y después de revisiones exhaustivas del ensayo, nos percatamos de la existencia de errores en los cálculos, que seguramente ocurrieron por azar, siendo en realidad cierta, en aquel escenario, la hipótesis alternativa. El error que se cometió en aquel entonces fue:

- 1) Un error tipo I.
- 2) Un error tipo II.
- 3) Un sesgo de selección.
- 4) Un sesgo de confusión.

RC: 2

En un ensayo clínico se comparan cifras de glucemia al final de un estudio con tres brazos de aleatorización, cada uno de ellos utilizando un tipo diferente de insulina (detemir frente a glargina frente a NPH). Si se quiere contrastar la hipótesis de la

diferencia entre las medias de los tres grupos, asumiendo que la variable glucemia se distribuye de forma normal en todos ellos, se utilizará:

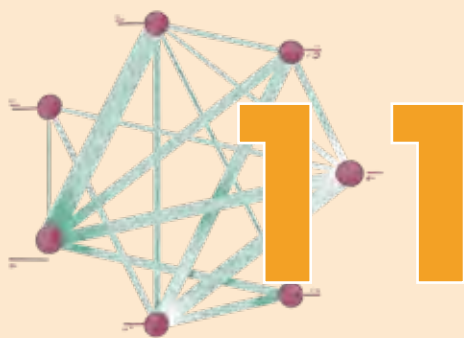
- 1) Coeficiente de correlación intraclass.
- 2) Coeficiente de correlación de Pearson.
- 3) Prueba de la t.
- 4) Análisis de la varianza.

RC: 4

En un ensayo clínico que estudió el riesgo de infarto agudo de miocardio asociado a fármacos se pretende valorar la influencia de la edad, el sexo y el hábito tabáquico en este desenlace. ¿Qué test estadístico emplearía para tal efecto?

- 1) Prueba de Student.
- 2) Correlación de Pearson.
- 3) Regresión lineal múltiple.
- 4) Regresión logística.

RC: 4



El tamaño muestral

Orientación MIR

Tema relativamente complejo y con preguntas en el MIR. Aquí suelen presentar un estudio tipo con varias posibilidades, de las cuales una no influye en el tamaño de la muestra.

Preguntas MIR

► MIR 13-14, 191

Conceptos clave

- Lo primero que se debe hacer cuando se pregunte sobre tamaño muestral, es conocer si se trata de una estimación de parámetros o de un contraste de hipótesis, para así saber qué es lo que se necesita en el estudio a realizar.
- Si se trata de un intervalo de confianza, entonces se debe conocer si se trata de una variable cualitativa (proporción) o de una cuantitativa (media). En ambos casos es preciso conocer el nivel de confianza y la precisión con la que se desea el estudio. Evidentemente, en el caso de la proporción, se necesitará conocer la proporción esperable y en el de las medias, la varianza esperable.
- En los casos de contraste de hipótesis, los aspectos más importantes a tener en cuenta son los errores α y β del estudio, la magnitud de la diferencia a detectar clínicamente relevante, la proporción de pacientes en los distintos grupos que responderán a cada tratamiento y la proporción de pérdidas.

11.1. Las muestras

Para realizar un estudio sobre una población, es necesario trabajar con una muestra, ya que abarcar a todos los miembros de una comunidad es prácticamente imposible. El **tamaño de la muestra** se determina en el diseño del estudio, de modo que sea apropiado para los objetivos del estudio y con los condicionamientos que se está dispuesto a asumir. Un número insuficiente de participantes impedirá encontrar las diferencias buscadas, concluyendo erróneamente que no existen, mientras que un número excesivo aumenta innecesariamente el coste.

Manejar muestras es más barato, rápido, fácil y suficientemente preciso y exhaustivo que manejar poblaciones. Los datos obtenidos en la muestra se generalizarán a la población original utilizando ciertas pruebas estadísticas con una mínima probabilidad de error (< 5% habitualmente). Hay dos campos en la inferencia estadística donde se va a desarrollar el cálculo del tamaño muestral: estimación de parámetros y contraste de hipótesis.

■ Estimación de parámetros

El tamaño dependerá de la variabilidad del parámetro que se desea estimar (esta ha de ser previamente conocida, o bien aproximarse a partir de datos preexistentes o estudios pilotos), de la precisión con que se pretenda obtener la estimación (amplitud deseada del IC, teniendo en cuenta que a mayor precisión, se necesitará un número más alto de sujetos) y del nivel de confianza. Para estimar una proporción se debe conocer:

- El nivel de confianza o seguridad ($1-\alpha$). Para una seguridad del 95% = 2; para una seguridad del 99% = 2,6.
- La precisión que se desea para el estudio.
- Una idea del valor aproximado del parámetro que se quiere medir, revisando la literatura o los estudios previos. En caso de no tener dicha información, se empleará el valor $p = 0,5$ (50%). En este caso, se necesita mayor n , pero garantizará una mejor estimación.

Por ejemplo, ¿a cuántas personas se tendría que estudiar para conocer la prevalencia de esquizofrenia en un área residencial de Madrid?

$$\text{Seguridad} = 95\%; \text{precisión} = 5\%; \text{proporción esperada} = \text{próxima al } 10\% \\ n = (2^2 \times p \times q) / d^2 = (4 \times 0,1 \times 0,9) / 0,05^2 = 144$$

Donde:

- 2^2 , ya que la seguridad es del 95%.
- p = proporción esperada (en este caso 10% = 0,1).
- $q = 1 - p$ (en este caso $1 - 0,1 = 0,9$).
- d = precisión (en este caso se desea un 5% = 0,05).

Para **estimar una media** se debe conocer:

- El nivel de confianza o seguridad ($1-\alpha$), habitualmente 95%.
- La precisión con que se desea estimar el parámetro.
- Una idea de la varianza de la variable cuantitativa que se supone existe en la población ► **MIR 13-14, 191**.

Por ejemplo, si se desea conocer la media de la colesterolemia basal de una población, con una seguridad del 95% y una precisión de ± 4 mg/dL, y se tiene información por un estudio piloto o revisión bibliográfica de que la varianza es de 300 mg/dL:

$$n = (2^2 \times 300) / 4^2 = 75$$

Recomenda

- La precisión está en el denominador y al cuadrado, lo que significa que cuanto más preciso se pretenda ser, más pequeño será el número del denominador y, por ello, el tamaño muestral requerido será más grande.
- En la estimación de parámetros, no se necesita conocer el error b para estimar el tamaño muestral.

■ Contraste de hipótesis

Se utiliza para estudios comparativos y, en él, el tamaño de la muestra indica el número aproximado de sujetos que se necesitan para detectar una diferencia determinada, si existe, con unos márgenes de error previamente definidos.

Los factores para determinar el tamaño muestral son:

- La magnitud de la diferencia a detectar que tenga interés clínicamente relevante. Se pueden comparar dos proporciones o dos medias.
- Tener una idea aproximada de los parámetros de la variable que se estudia (bibliografía, estudios previos).
- La variabilidad de la variable principal.
- La seguridad del estudio (riesgo de cometer un error α : 5%).
- El poder estadístico ($1-\beta$: 80% al menos) (riesgo de cometer un error β : 0,1-0,2).
- La proporción de pacientes en los distintos grupos que responderá a cada tratamiento.
- La proporción de pérdidas (d): se deberá multiplicar el tamaño calculado por $1 / (1-d)$.
- La definición de si la hipótesis va a ser unilateral o bilateral:
 - **Bilateral**. Cualquiera de los dos parámetros a comparar (medias o proporciones) puede ser mayor o menor que el otro. No se establece dirección. H_0 diferente de H_1 .
 - **Unilateral**. Cuando se considera que uno de los parámetros debe ser mayor que el otro, indicando, por tanto, una dirección de las diferencias. Solo $H_0 > 0 < H_1$.

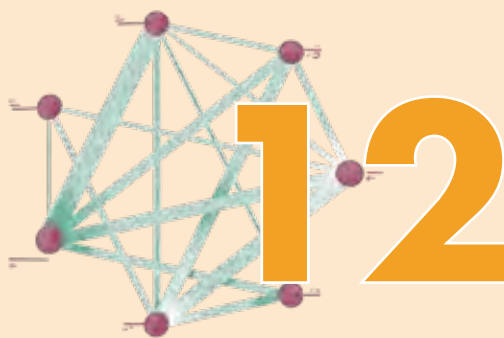
La hipótesis bilateral es una hipótesis más conservadora y disminuye el riesgo de cometer un error de tipo I.

Casos clínicos

Se quiere realizar un estudio para estimar la prevalencia de diabetes mellitus en la población joven de Barcelona. Para determinar el número de sujetos necesario para el estudio, se tendrá en cuenta todo lo siguiente, excepto:

- 1) La prevalencia estimada de diabetes mellitus en la población.
- 2) El nivel de precisión que se desea para la estimación.
- 3) La variabilidad de la diabetes mellitus en la población.
- 4) La potencia necesaria.

RC: 4



La salud y los sistemas sanitarios

Orientación MIR

Tema con pocas preguntas en el MIR. Destacan las definiciones del primer subapartado, especialmente los determinantes de la salud y la promoción y prevención de la misma, así como las características del SNS español.

Lo más importante de la planificación sanitaria es conocer en que consiste y los tipos de indicadores. Además, trata de entender las diferencias generales entre niveles de planificación.

Preguntas MIR

► MIR 17-18, 218

Conceptos clave

- ◉ El determinante de la salud más relevante es el conjunto de los hábitos de vida.
- ◉ El determinante en el que más se invierte es el sistema de cuidados.
- ◉ El sistema de salud se compone de sistema sanitario y de otros elementos, como las redes de cuidados informales.
- ◉ El Sistema Nacional de Salud (SNS) español es un ejemplo de modelo Beveridge de servicio de salud.
- ◉ El SNS garantiza financiación pública, universalidad, equidad, gratuidad y atención integral.
- ◉ La planificación sanitaria incluye la planificación de la salud y la planificación de la atención a la salud.
- ◉ Se debe priorizar por importancia y/o por vulnerabilidad del problema de salud.
- ◉ Una necesidad de salud es la diferencia entre el estado de salud observado (problema) y el estado de salud deseado.
- ◉ La planificación normativa o política desarrolla objetivos abstractos e ideológicos a largo plazo.
- ◉ La planificación estratégica produce planes de salud (hojas de ruta) más concretas.
- ◉ La planificación táctica y la operativa establecen e implementan programas de salud específicos.
- ◉ La evaluación es común a todo el proceso de planificación y emplea indicadores, criterios y estándares o normas.

12.1. Introducción

La Organización Mundial de la Salud (OMS) llama **salud** al estado completo de bienestar físico, mental y social.

De acuerdo con el modelo de Lalonde, existen cuatro **factores determinantes de la salud**:

- Biológicos o genéticos.
- Ambientales.
- Secundarios a hábitos de vida.
- Dependientes del sistema de cuidados.

Estos factores se relacionan entre sí mediante interacciones muy complejas. Algunos autores, como Vicente Pastor, afirman que son el ambiente y los hábitos los que más determinan la salud de las poblaciones (**Figura 12.1**).

De los cuatro factores, es el cambio en los hábitos de vida el que produce mayor ganancia en salud o reducción en morbilidad.

Se llama **promoción de la salud** al proceso que permite a las personas incrementar el control de su salud para mejorarla, incidiendo de manera individual. Una de las estrategias principales para ello es la **educación para la salud**, que consiste en la combinación de actividades de información y educación que llevan a que la población desee estar sana, sepa alcanzar su salud y haga lo que pueda a nivel individual o colectivo para mantener su salud.

La **protección de la salud** incluye las acciones que mejoran la salud, incidiendo en factores ambientales como la salud ambiental y la seguridad alimentaria.

Recuerda

- El cambio en los hábitos de vida produce la mayor ganancia en salud, mientras que el mayor gasto está dirigido a los sistemas de cuidados.

12.2. Los sistemas sanitarios

El término *sistema de salud* engloba todas las actividades orientadas a promover, mantener o restablecer la salud de la población. El **sistema sanitario (SS)** es uno de sus componentes. Incluye el conjunto de organizaciones, instituciones y recursos formales que se emplean con tres objetivos finales:

- Mejorar la salud individual y colectiva.
- Ofrecer un trato adecuado a los usuarios.
- Garantizar su propia seguridad financiera.

Otro componente relevante del sistema de salud son las **redes informales de cuidado** (familiares, vecinos), que suponen hasta el 88% de la respuesta social a la enfermedad.

Cada SS tiene cuatro elementos o **actores fundamentales**:

- Usuarios.
- Proveedores de atención sanitaria.
- Financiadores.
- Administración.

Y se estructura en torno a tres **funciones**:

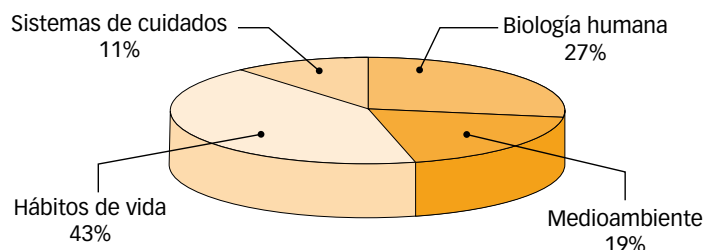
- Financiación del sistema.
- Prestación de los servicios médicos.
- Regulación normativa.

En la práctica, existen tres modelos clásicos de SS basados en servicio de salud, seguridad social o libre mercado sanitario (**Tabla 12.1**). Sin embargo, la mayor parte de sistemas tienen peculiaridades propias y en todos los casos coexisten servicios públicos y privados.

	SERVICIO DE SALUD (BEVERIDGE)	SEGUROS SOCIALES (BISMARCK)	LIBRE MERCADO SANITARIO (LIBERAL)
Financiación	Pública (impuestos generales)	Pública (cotizaciones)	Privada (directa o cotizaciones)
Prestación de servicios médicos	Pública con cobertura universal	Pública/privada	Privada
Ventajas	Equidad y control del gasto	Libertad de elección	Competitividad: estimula avances científicos

Tabla 12.1. Modelos de sistema sanitario.

INFLUENCIA DE CADA DETERMINANTE



INVERSIÓN EN CADA DETERMINANTE

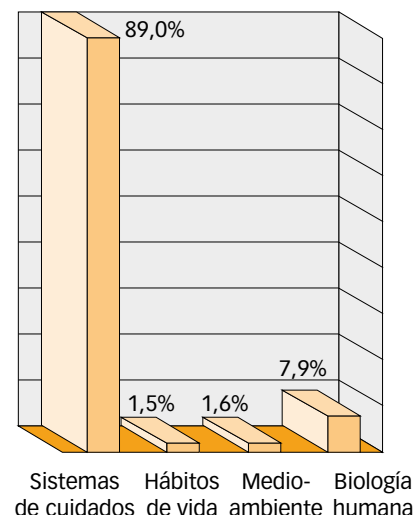


Figura 12.1. Determinantes de la salud por influencia y por inversión (modificado de Lalonde, 1974).



12.3. El Sistema Nacional de Salud español

El actual **Sistema Nacional de Salud (SNS)** de España es un ejemplo de modelo de Beveridge, y se regula principalmente por la **Ley 14/1986 General de Sanidad**.

El SNS se caracteriza por ► **MIR 17-18, 218** :

- Financiación pública.
- Universalidad.
- Gratuidad.
- Atención integrada (integración de promoción, prevención, diagnóstico, tratamiento y rehabilitación).
- Atención integral (esferas física, psíquica y social).
- Descentralización territorial (en 2002 se completó la transferencia de los servicios sanitarios a las comunidades autónomas).

La **Ley 16/2003 de Cohesión y Calidad del SNS** y el Real Decreto 1030/2006 definen el catálogo de prestaciones para minimizar las desigualdades y garantizar la equidad en el territorio nacional. Se llama **equidad**, en este contexto, a la posibilidad que tiene un sujeto de ser atendido por el sistema sanitario, independientemente de su condición social, sexo o lugar de nacimiento.

El **Consejo Interterritorial del Sistema Nacional de Salud**, organismo encargado de coordinar las diferentes Consejerías de Salud autonómicas, tiene un carácter consultor, y sus recomendaciones no son de obligado cumplimiento. Como resultado, en este momento existen múltiples fórmulas de gestión de los servicios sanitarios dentro del SNS español, con modalidades públicas, privadas y mixtas.

En 2002 se completó la **transferencia** de los servicios sanitarios a las comunidades autónomas, con lo que se **descentralizó** la administración del SNS.

12.4. La planificación sanitaria

La **planificación sanitaria** permite escoger, organizar y evaluar los recursos y actividades necesarios para satisfacer las necesidades de salud de una población. Incluye la **planificación de la salud** (la modificación de los determinantes de la salud) y la planificación de la atención de la salud, siendo esta última la principal receptora de inversión.

En la planificación sanitaria se suceden varias funciones:

1. **Identificación de la necesidad de salud.** Es la diferencia entre el estado de salud observado (problema) y el estado de salud deseado. Puede partir de las recomendaciones de expertos o de lo percibido por la población. Se utilizan indicadores de salud, encuestas de salud o metodología cualitativa.
2. **Establecimiento de prioridades.** Según **importancia** (extensión y gravedad) y **vulnerabilidad** del problema (posibilidad potencial de solucionarlo con una intervención).
3. **Formulación de objetivos.**
4. **Identificación de recursos disponibles.**
5. **Determinación de actividades a implementar.**
6. **Ejecución del programa.**
7. **Evaluación.**

Recuerda

- Se debe priorizar según la importancia del problema y/o según la vulnerabilidad del mismo a ser solucionado.

Niveles de planificación

Existen diferentes tipos de planificación según los objetivos perseguidos, las herramientas empleadas y el período de tiempo de implementación.

Planificación normativa o política

¿La planificación normativa se formula la siguiente pregunta: ¿qué se quiere conseguir? Desarrolla objetivos globales a largo plazo, establecidos en un ámbito local, nacional o multinacional (OMS, Banco Mundial, UE). El resultado es la **política de salud**. Proporciona el marco en el que se desarrollan los demás niveles de planificación.

Planificación estratégica

La planificación estratégica se formula la pregunta: ¿cómo se puede conseguir? Su finalidad es trasladar a la práctica las líneas políticas: establecer una hoja de ruta. Persigue objetivos a medio plazo y selecciona las prioridades de acción de acuerdo con la política de salud. El resultado es el plan de salud.

Las planificaciones normativa y estratégica deben cumplir los criterios de **pertinencia** (justificación de la definición del objetivo a abordar) y **suficiencia** (grado de definición concreta del problema de salud).

Planificación táctica o estructural

La planificación táctica o estructural realiza el diagnóstico comunitario de las necesidades de salud, determina concretamente las actividades a realizar para lograr el objetivo estratégico y distribuye los recursos.

Planificación operativa

La planificación operativa ejecuta los objetivos tácticos mediante objetivos operacionales, a corto plazo.

El resultado de las planificaciones táctica y operativa es el diseño e implementación de un **programa de salud**, el conjunto actividades dirigidas a obtener esos objetivos específicos. Todas ellas se integran entre sí jerárquicamente (**Tabla 12.2**).

Evaluación

La **evaluación** es el proceso de medición de situaciones y su comparación con el objetivo. Nace de la necesidad de optimizar los recursos limitados. Es un componente común a todos los niveles de planificación sanitaria.

Utiliza indicadores, criterios y estándares.

- **Indicadores.** Variables que cuantifican una dimensión conceptual y producen un número. Permiten el establecimiento de comparaciones.

PLANIFICACIÓN	GENERALIDADES	RESULTADO	EJEMPLO
Normativa	Objetivos políticos, ideológicos Ámbito estatal, europeo, mundial	Políticas de salud	Mejorar la salud de los pacientes con trastorno mental grave
Estratégica	Hoja de ruta para desarrollar la planificación normativa: • Identifica problemas • Establece prioridades	Planes de salud	Aumentar la inserción laboral de los pacientes con trastorno mental grave
Táctica	Diagnóstico de causas y desarrollo de actividades para resolver los problemas identificados por la planificación estratégica: • Objetivos específicos • Diseño de actividades • Previsión de recursos	Programas de salud	Desarrollar una red de rehabilitación laboral específica para los enfermos con trastorno mental grave (calculando su coste e impacto potencial y diseñando recursos a implementar)
Operativa	Fija objetivos operacionales a través de los que ejecutar el programa de salud		Lograr una inserción del 50% de los pacientes con trastorno mental grave

Tabla 12.2. Niveles de planificación sanitaria.

Los indicadores pueden ser:

- De estructura. Si los recursos asignados a un programa son suficientes en cantidad o calidad.
- De proceso. Si la actividad ha sido cumplida y la calidad del cumplimiento.
- De resultado. Si la actividad ha generado los objetivos previstos.
- **Criterios.** Conjuntos de reglas que hay que cumplir para afirmar que una estructura o un proceso se han desarrollado de forma adecuada.
- **Estándar o norma.** Nivel que debe alcanzar un criterio para considerarse óptimo.

Recuerda

- La evaluación es un paso común a todos los niveles de planificación.
- Los indicadores son variables numéricas y pueden cuantificar estructura, proceso o resultado.

La evaluación es común a todo el proceso de planificación (**Figura 12.2**).

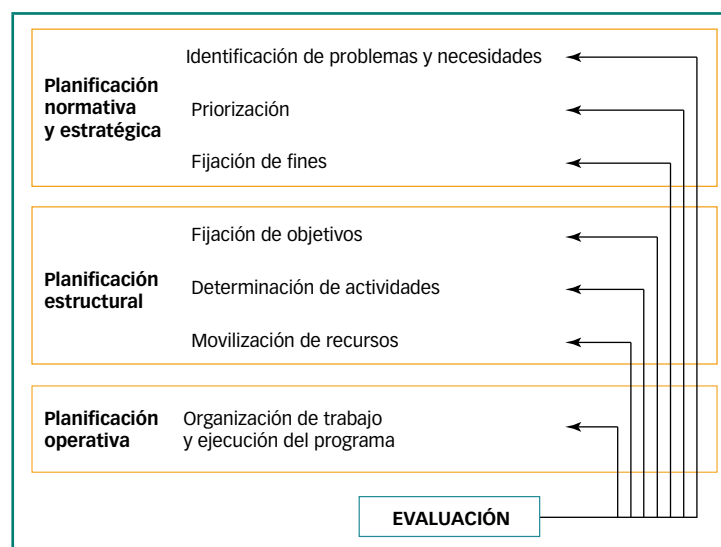
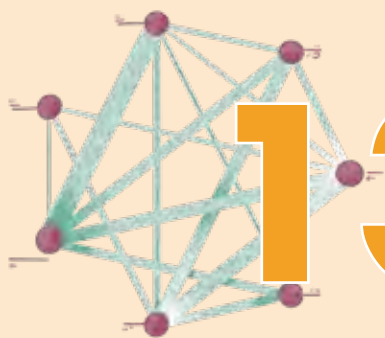


Figura 12.2. Papel de la evaluación en el proceso de planificación sanitaria.



La atención sanitaria

Orientación MIR

Tema de cierta importancia. Recuerda las funciones principales de la Atención Primaria, el concepto de grupos relacionados por el diagnóstico (GRD), así como el de la cirugía mayor ambulatoria (CMA). Algunos indicadores hospitalarios y el conjunto mínimo básico de datos (CMBD) han sido preguntados de forma marginal en alguna ocasión.

Preguntas MIR

- MIR 19-20, 154
- MIR 18-19, 125; MIR 18-19, 206
- MIR 16-17, 127
- MIR 15-16, 171

Conceptos clave

- ◉ La atención primaria se organiza en áreas de salud (200.000-250.000 habitantes) que se dividen en zonas básicas de salud (5.000-25.000 habitantes), llevándose a cabo la actividad asistencial en los centros de salud.
- ◉ Atención integral en atención primaria significa abordar las esferas física, psíquica y social.
- ◉ Atención integrada en planificación significa integrar promoción, protección, prevención, diagnóstico, tratamiento, rehabilitación y reinserción social.
- ◉ Los objetivos de la prevención a distintos niveles son: evitar la aparición de la enfermedad (primaria), modificar la historia natural de la enfermedad (secundaria), mejorar la calidad de vida de los enfermos (terciaria) y disminuir la iatrogenia innecesaria (cuaternaria).
- ◉ El mayor gasto sanitario va dirigido a los sueldos de los trabajadores de la atención especializada.
- ◉ Los datos obligados en el informe de alta hospitalario, aprobados por el Consejo Interterritorial del Ministerio de Sanidad y Consumo, se basan en el conjunto mínimo básico de datos (CMBD) al alta hospitalaria.
- ◉ Los grupos relacionados por el diagnóstico (GRD) son el sistema de clasificación de los pacientes hospitalizados más utilizado y lo hacen en función del isoconsumo de recursos (sobre todo) y por mismo proceso clínico.
- ◉ La cirugía mayor ambulatoria (CMA) consiste en procedimientos quirúrgicos programados, con anestesia local, locorregional o general, que no requieren ingreso hospitalario, por necesitar cuidados posoperatorios poco intensivos.
- ◉ La CMA constituye un programa muy eficiente que además mejora la calidad extrínseca o percibida por los pacientes.

La **atención sanitaria** en España puede dividirse en atención primaria y en atención hospitalaria o especializada. Ambas deben estar coordinadas para conseguir unos programas de salud efectivos y eficientes.

13.1. Atención primaria

La **atención primaria** es la puerta de entrada al sistema sanitario y, por ello, constituye una pieza clave en el correcto funcionamiento de este, así como un punto imprescindible para la planificación sanitaria. Tiene el papel de actuar como filtro al acceso a la atención hospitalaria de cara a optimizar el consumo de recursos, lo que se conoce con el nombre de *gate-keeping*. Además, fomenta la relación y confianza médico-paciente dada la continuidad asistencial que comporta.

La declaración de Alma-Ata, aprobada por la OMS, sitúa la atención primaria como estrategia para alcanzar el mejor nivel de salud de los pueblos. La define como una atención sanitaria basada en técnicas sencillas, científicamente fundamentadas y socialmente aceptables, al alcance de los individuos, familia y comunidad a un coste aceptable para la comunidad y el país. Otorga un papel activo a la ciudadanía, reclama un derecho a la salud sin discriminaciones (equidad) y responsabiliza a los gobiernos de garantizar el cuidado de la salud de sus individuos ► **MIR 18-19, 125**.

Las **funciones principales** de la atención primaria consisten en:

- **Asistencia sanitaria** a demanda, programada y urgente, tanto en consulta como en domicilio (articulación asistencial) a lo largo de la vida del enfermo (atención longitudinal).
- **Indicación y prescripción** de exploraciones complementarias y tratamientos. En caso de existir varias alternativas terapéuticas intercambiables, el sistema de precios de referencia determina el precio a pagar por ese grupo de medicamentos equivalentes ► **MIR 16-17, 127**.
- **Atención al individuo, la familia y la comunidad** de forma integral desde la atención primaria, esto es, basándose no solo en la esfera física, sino también en la esfera psíquica y en la esfera social ► **MIR 15-16, 171**.
- Aportar una **planificación sanitaria** integrada en relación a programas dirigidos a patologías prevalentes con relevancia social: promoción y protección de la salud, prevención, diagnóstico, tratamiento, rehabilitación y reinserción social.
- Realización de **programas específicos**: educación de la salud, salud pública, salud mental, atención a drogodependencia, cuidados paliativos, dependencia en la tercera edad, atención a la mujer y violencia de género, salud bucodental...

Recuerda

- Atención integrada en planificación sanitaria no es sinónimo de atención integral en atención primaria.

La **prevención de la salud** se define como el conjunto de medidas necesarias para evitar el desarrollo o progreso de enfermedades. La prevención se realiza a todos los niveles de la atención sanitaria, si bien la Atención Primaria es donde tiene más importancia. Existen cuatro tipos de prevención en función de su objetivo:

- **Primaria**: evitar la aparición de enfermedad (p. ej., vacunas).

- **Secundaria**: modificar la historia natural de la enfermedad detectándola en estadios asintomáticos e iniciando tratamiento precoz (p. ej., cribado mediante mamografía).
- **Terciaria**: mejorar la calidad y esperanza de vida de los pacientes (p. ej., rehabilitación cardíaca tras infarto agudo de miocardio).
- **Cuaternaria** ► **MIR 19-20, 154; MIR 18-19, 206**: disminuir la iatrogenia evitando las intervenciones médicas innecesarias (p. ej., no realizar una angio-TC a un paciente con enfermedad renal crónica grave si se puede diagnosticar mediante otra técnica de imagen).

Recuerda

- El cribado o *screening* es una estrategia de prevención secundaria.

Las competencias de la gestión sanitaria y los presupuestos en cada ámbito territorial dependen de cada comunidad autónoma. Son coordinadas, de forma consultiva, por el Consejo Interterritorial del Sistema Nacional de Salud (SNS).

La atención primaria en cada comunidad autónoma se organiza en áreas de salud, delimitadas por factores geográficos, demográficos, socioeconómicos, epidemiológicos, climatológicos y de dotación de instalaciones y medios de comunicación. Cada área de salud da cobertura a 200.000-250.000 habitantes y debe elaborar un plan de salud que incluya las actividades sanitarias necesarias para cumplir los objetivos.

Un área de salud se compone de varias **zonas básicas de salud**, también delimitadas por criterios demográficos, geográficos y sociales, para dar servicio a 5.000-25.000 habitantes por zona básica de salud. Estas zonas básicas de salud tienen estructuras físicas y funcionales, los **centros de salud**, donde se desarrollan las actividades del equipo de atención primaria. Los centros de salud deben estar a un máximo de 30 minutos, en transporte privado, de cualquier municipio perteneciente a la zona básica de salud. Sus funciones son:

- Albergar la estructura física para consultas y servicios asistenciales, así como los recursos para realización de exploraciones complementarias.
- Facilitar el trabajo en equipo de los profesionales sanitarios y no sanitarios (médicos de medicina familiar y comunitaria, pediatras, diplomados en enfermería, auxiliares de enfermería, matronas, fisioterapeutas, odontólogos, trabajadores sociales, personal de admisión y mantenimiento...).
- Servir de centro de reunión entre la comunidad y los profesionales sanitarios.
- Mejorar la organización administrativa.

13.2. Atención hospitalaria o especializada

La **atención hospitalaria o especializada** es la que más gasto sanitario consume; el apartado más importante de este es el sueldo del personal sanitario. El objetivo de la atención especializada es ofrecer los servicios no disponibles desde atención primaria; incluye el ingreso hospitalario o a domicilio, así como urgencias y emergencias graves; mantiene objetivos comunes con la atención primaria como serían la formación de profesionales y el fomentar la promoción de la salud, la prevención y la educación sanitaria.



Los hospitales pueden clasificarse de distintas formas (Tabla 13.1).

	TIPOS DE HOSPITALES
Por función	Generales, especiales y monográficos
Por estancia	Corta estancia o de agudos Larga estancia o de crónicos
Por complejidad	Baja: medicina interna, cirugía, ginecología, pediatría Media: diversas especialidades médicas y quirúrgicas Alta: subespecialidades
Por dependencia patrimonial	Públicos Privados (con o sin ánimo de lucro)
Por demarcación territorial	Hospital de área Hospital de referencia de área Hospital de referencia de la comunidad autónoma CSUR (centros, servicios y unidades de referencia)

Tabla 13.1. Tipos de hospitales.

Es necesario recabar un conocimiento útil y ajustado a la realidad de la atención especializada, con el objetivo de evaluar estructuras, procesos y resultados, para realizar mejoras del sistema sanitario. Los sistemas de información sanitaria (SIS) se encargan de la recogida, procesamiento, análisis y tratamiento de la información. Existen dos SIS principales en España, aprobados por el Consejo Interterritorial del SNS (Tabla 13.2):

- Encuesta de morbilidad hospitalaria (EMH).
- Conjunto mínimo básico de datos (CMBD) o *Minimum Data Set* (MDS).

SIS	DATOS
EMH	Edad, sexo, estado civil, lugar de residencia, diagnóstico al alta (previamente consideraba el diagnóstico al ingreso), tiempo de estancia y provincia de la hospitalización
CMBD	Identificación del paciente, fecha de nacimiento, sexo, residencia, identificación del hospital (no el tipo de hospital), financiación del hospital, número de historia clínica, número de asistencia o episodio, fecha de ingreso, circunstancia del ingreso (programada, traslado, desde urgencias...), servicio al ingreso, servicio al alta, diagnóstico principal y otros diagnósticos (codificados), procedimientos quirúrgicos/obstétricos y otros procedimientos (codificados), fecha del alta, circunstancias del alta (a domicilio, traslado, alta voluntaria, <i>exitus</i> ...)

EMH: encuesta de morbilidad hospitalaria; CMBD: conjunto mínimo básico de datos.

Tabla 13.2. Tipos de sistemas de información sanitaria (SIS).

■ Indicadores hospitalarios y sistemas de clasificación de pacientes

La **estancia media** es el factor limitante más importante para el ingreso de los pacientes. Además de esta, cabe destacar el **índice de ocupación (IO)**, que idealmente debería superar el 85%, y el **número de camas necesarias (N.º camas)**, que suele expresarse por cada 1.000 habitantes. Para calcularlo es necesario conocer el concepto de **frecuentación hospitalaria**, que consiste en el número de ingresos por cada 1.000 habitantes-tiempo en una población determinada.

$$\text{Estancia media} = \frac{\text{N.º días de estancia en el período}}{\text{N.º episodios de hospitalización en el período}}$$

$$\text{Índice de ocupación} = \frac{\text{N.º días de estancia en el período}}{\text{N.º camas} \times \text{N.º días del período}}$$

$$\text{N.º camas} = \frac{\text{N.º ingresos} \times \text{Estancia media}}{\text{Índice de ocupación} \times \text{N.º días del período}}$$

$$\text{N.º ingresos} = \text{Población} \times \text{Frecuentación hospitalaria}$$

Ejemplo. Supóngase que una población de 150.000 habitantes tiene una frecuentación hospitalaria de 120 ingresos/1.000 habitantes-año. ¿Cuántas camas se necesitarían en un mes si la estancia media es de 8 días y el índice de ocupación esperado es del 90%? ¿A cuántas camas por cada 1.000 habitantes equivaldrían?

$$\text{N.º camas} = \frac{1.500 \times 8}{0,9 \times 30} = 445$$

Nótese que el número de ingresos (1.500) viene dado de multiplicar la población (150.000 habitantes) por la frecuentación hospitalaria mensual (0,01 = 10 ingresos/1.000 habitantes-mes). Esto es así porque se pide el número de camas en un mes. Si lo pidieran en un año, no sería necesario realizar la equivalencia (en ese caso la frecuentación hospitalaria sería 0,12 = 120 ingresos/1.000 habitantes-año).

$$\text{N.º camas/1.000 habitantes} = \frac{445}{150.000} \times 1.000 = 2,96$$

Existen otros indicadores hospitalarios, entre los que destacan los que se enumeran en la Tabla 13.3.

Los **sistemas de clasificación de pacientes** tienen por objetivo conseguir información para comparar indicadores hospitalarios ajustados, especialmente la estancia media y los costes, entre distintos centros. Esto es así porque no es correcto comparar los resultados y el consumo de recursos globales entre centros, dada la diversidad de la complejidad y gravedad de las patologías que motivan el ingreso en función del tipo de hospital.

INDICADOR	FÓRMULA
Índice de rotación paciente/cama	$\frac{\text{Ingresos en el período}}{\text{N.º camas disponibles en el período}}$
Intervalo de sustitución	$\frac{\text{Estancia media} \times (1 - \text{índice ocupación})}{\text{Índice ocupación}}$
Presión de urgencias	$\frac{\text{N.º ingresos desde urgencias}}{\text{N.º ingresos totales}}$
Rendimiento de quirófano	$\frac{\text{Horas de quirófano utilizado}}{\text{Horas de quirófano disponibles}}$
Porcentaje de sustitución	$\frac{\text{N.º cirugía mayor ambulatoria}}{\text{N.º cirugías totales}}$
Impacto estancias ahorradas	$\text{N.º cirugía mayor ambulatoria} \times \text{estancia media}$
Índice de consultas	$\frac{\text{N.º visitas sucesivas}}{\text{N.º primeras visitas}}$

Tabla 13.3. Indicadores hospitalarios.

Existen varios sistemas de clasificación de pacientes:

- **Hospitales de agudos.** GRD o GDR (grupos relacionados por el diagnóstico o grupos de diagnósticos relacionados). Tienen dos variantes: AP-DRG (*All Patients Diagnosis Related Groups*) y RDRG (*Refined Diagnosis Related Groups*).
- **Unidades de cuidados intensivos.** APACHE (*Acute Physiology Age Chronic Health Evaluation*), MPM (*Mortality Probability Model*), PRISM (*Pediatric Risk Score of Mortality*) y TISS (*Therapeutic Intervention Scoring System*).
- **Centros de larga estancia.** RUG (*Ressource Utilization Groups*).

Los **grupos relacionados por el diagnóstico (GRD)** son el sistema de clasificación de pacientes más utilizado en gestión hospitalaria, no solo en España, sino en la mayoría de países. Clasifica los pacientes hospitalizados por isoconsumo de recursos (sobre todo) y por mismo proceso clínico. Intentan comparar centros de acuerdo con el coste asociado al completar un alta hospitalaria, lo que se utiliza en modelos de sistema sanitario de gestión privada, como ocurre en Estados Unidos.

Mediante los GRD se puede definir dos indicadores hospitalarios de comparación:

- Índice de **estancia media ajustada por funcionamiento (EMf)**. Un EMf > 1 indica que el hospital presenta una estancia media más larga para un nivel de complejidad determinado.
- Índice de **estancia media ajustada por casuística (EMc)**. Un EMc > 1 indica que el hospital atiende a pacientes de mayor complejidad que el promedio de hospitales de comparación.

Recuerda

- Los GRD clasifican a los pacientes hospitalizados por isoconsumo de recursos (sobre todo) y por mismo proceso clínico.

Cirugía mayor ambulatoria

Se consideran **cirugía mayor ambulatoria (CMA)** aquellos procedimientos quirúrgicos programados, realizados mediante anestesia general, locorregional o local con sedación, que no requieren ingreso hospitalario, por

necesitar cuidados posoperatorios poco intensivos, pudiendo ser dados de alta el mismo día de la intervención.

Para llevar a cabo un programa de CMA, se requiere una unidad multidisciplinar y se debe asegurar que la seguridad de los procedimientos quirúrgicos (complicaciones, tasa de reintervención...) tiene las mismas garantías que en régimen de hospitalización.

Recuerda

- La cirugía mayor ambulatoria es programada, puede utilizar cualquier tipo de anestesia y no requiere ingreso hospitalario.

Existe también el concepto de **cirugía menor ambulatoria** que incluye aquellas intervenciones realizadas con anestesia local sin sedación y que permiten el alta inmediata.

Ventajas de la CMA:

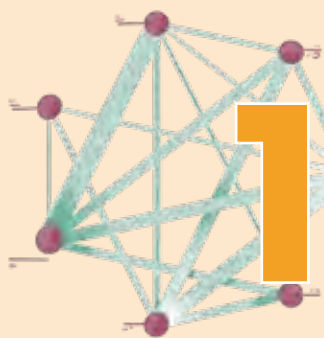
- Menor estrés psicológico, ansiedad y alteración de la rutina del paciente.
- Recuperación más rápida.
- Menor riesgo de infección nosocomial.
- Mayor eficiencia: reducción del coste, la estancia hospitalaria y la lista de espera quirúrgica (aumento de la calidad percibida).
- Fomenta la interacción entre atención primaria y atención especializada.

Desventajas de la CMA:

- El paciente y los familiares juegan un papel más importante en el posoperatorio (curas de la herida, dietas específicas, convalecencia...).
- Puede beneficiar a las clases sociales medias-altas ya que, en ocasiones, los factores sociales son un condicionante para su realización (necesidad de cuidados).

Los criterios de selección para la realización de una CMA dependen del:

- **Paciente.** Riesgo quirúrgico bajo (ASA I o II), ausencia de trastornos de la coagulación, ausencia de patología psiquiátrica, apoyo familiar.
- **Procedimiento.** Bajo riesgo de sangrado, duración inferior a 90 minutos, dolor controlable con analgésicos orales, no apertura de cavidad torácica o abdominal (a excepción de la vía laparoscópica).



La evaluación económica

Orientación MIR

Tema más relevante de Planificación y Gestión, y uno de los más importante de toda la preparación MIR, tanto por el elevado y constante número de preguntas que incluye, como por su rentabilidad. Recuerda las definiciones de eficacia, efectividad y eficiencia, así como los tipos de coste. En cuanto a los análisis de costes, es de interés conocer la perspectiva desde la que están enfocados, las unidades de medida de los resultados, cuándo utilizar un análisis de minimización de costes y entender el concepto de análisis incremental.

Preguntas MIR

- ▷ MIR 21-22, 49
- ▷ MIR 20-21, 56
- ▷ MIR 19-20, 183
- ▷ MIR 18-19, 20; MIR 18-19, 124
- ▷ MIR 17-18, 216; MIR 17-18, 217
- ▷ MIR 16-17, 125; MIR 16-17, 126
- ▷ MIR 15-16, 195; MIR 15-16, 196
- ▷ MIR 14-15, 187; MIR 14-15, 188
- ▷ MIR 13-14, 208

Conceptos clave

- ⊙ La eficacia y la efectividad miden los resultados en condiciones ideales y reales, respectivamente.
- ⊙ La eficiencia relaciona los resultados con el coste invertido, por lo que una alternativa más eficaz/efectiva puede no ser la más eficiente.
- ⊙ Los costes sanitarios directos son todos aquellos que tienen relación con la asistencia médica.
- ⊙ Los costes indirectos (no sanitarios) son aquellos derivados de la disminución de la productividad laboral secundaria a patología médica.
- ⊙ Las evaluaciones económicas suelen realizarse desde la perspectiva de la sociedad en general, e incluyen los costes directos e indirectos.
- ⊙ Los resultados se miden mediante unidades monetarias en el análisis coste-beneficio, unidades físicas (años de vida ganados) en el análisis coste-efectividad y unidades de utilidad (años de vida ajustados por calidad, AVAC) en el análisis coste-utilidad.
- ⊙ Ante dos alternativas con resultados similares, el análisis de minimización de costes informa de la más eficiente, por ser la más económica. Dicho análisis se utiliza para comparar genéricos con el fármaco de referencia.
- ⊙ El análisis incremental es de elección en los análisis coste-efectividad y coste-utilidad y consiste en comprobar si una alternativa es coste-efectiva o coste-útil respecto a otra en función de un umbral fijado *a priori*.

Evaluar consiste en medir una situación y compararla con el objetivo fijado *a priori*. Extrapolado a la planificación, consiste en analizar la diferencias entre lo planificado, lo ejecutado y los resultados obtenidos, teniendo en cuenta los recursos empleados. El objetivo de la evaluación es proponer medidas correctoras que mejoren los resultados, dada la necesidad de optimización de los recursos por su escasez. La evaluación es necesaria en todas las etapas de la planificación, siendo la evaluación económica de especial importancia en la planificación táctica.

14.1. Eficacia, efectividad y eficiencia

Los resultados de un programa de salud pueden definirse en función de su (Tabla 14.1) ► MIR 20-21, 56; MIR 15-16, 196 :

- **Eficacia.** Resultados obtenidos en **condiciones ideales** (eficaz en el matraz). Responde a la pregunta: ¿puede funcionar en condiciones ideales? Es la que se obtiene en la mayoría de los ensayos clínicos. También recibe el nombre de eficacia potencial, por corresponder a los máximos resultados que puede aportar una intervención.
- **Efectividad.** Resultados obtenidos en **condiciones reales** (efectivo *in vivo*). Responde a la pregunta: ¿puede funcionar en la práctica clínica habitual? Es la que se obtiene en los estudios observacionales. Un sinónimo es el de eficacia real, por corresponder a los resultados obtenidos en la población general fuera de un ensayo clínico.
- **Eficiencia.** Resultados obtenidos en relación a los **recursos económicos invertidos**, es decir, constituye el coste/efectividad (eficiente el gerente). Responde a la pregunta: ¿es rentable desde el punto de vista económico?

CONCEPTO	RESULTADOS OBTENIDOS EN	RESPONDE A
Eficacia	Condiciones ideales	¿Puede funcionar en condiciones ideales?
Efectividad	Condiciones reales	¿Puede funcionar en la práctica clínica habitual?
Eficiencia	Relación a los recursos económicos invertidos	¿Es rentable desde el punto de vista económico?

Tabla 14.1. Conceptos de eficacia, efectividad y eficiencia.

Recuerda

- Regla mnemotécnica:
Eficaz en el matraz. Efectivo *in vivo*. Eficiente el gerente.

Los resultados de los ensayos clínicos, en general, informan de la eficacia de una intervención. Esto es así por las condiciones especiales de selección, el estrecho seguimiento y por el cumplimiento que se exige a los participantes.

En muchos estudios, especialmente aquellos más explicativos, existe un período de cumplimiento del tratamiento pre-aleatorización, tras el cual se excluyen del estudio los participantes no cumplidores: no es posible esperar que los resultados obtenidos se mantengan en la práctica clínica. Para conocer la efectividad de una intervención son necesarios ensayos clínicos pragmáticos. Otra alternativa son los estudios observacionales (cohorte y casos-contróles), que evalúan a los participantes en condiciones reales. Sin embargo, ofrecen evidencia científica de peor calidad que los ensayos clínicos.

Una buena eficacia no siempre implica una efectividad satisfactoria. Por ejemplo, bien por motivos económicos (imposibilidad de pagar el precio de un fármaco en países con sanidad exclusivamente privada) o por incumplimiento de las medidas higienicodietéticas (pacientes hipertensos que no cumplen una dieta hiposódica), es posible que un fármaco que ha demostrado eficacia en un ensayo clínico no sea efectivo en una población determinada. Por otro lado, una vez comprobada la efectividad de una intervención en una población, se asume que se mantendrán los resultados cada vez que se implemente, y no habrá que volver a comprobarla.

Finalmente, una intervención más eficaz/efectiva que otra puede ser menos eficiente si el coste es proporcionalmente mayor al beneficio. Por ejemplo, supóngase que el fármaco A prolonga la esperanza de vida 5 años de media, con un coste total de 5.000 €, mientras que el fármaco B prolonga la esperanza de vida 6 años de media con un coste total de 15.000 €; el fármaco A es menos eficaz/efectivo (5 años *versus* 6 años) pero es más eficiente (1.000 €/año *versus* 2.500 €/año).

14.2. Tipos de costes

Los costes pueden clasificarse en ► MIR 14-15, 187 :

- **Costes directos sanitarios.** Relacionados con la asistencia médica. Pueden ser fijos o estructurales, cuando no dependen de la cantidad de actividad asistencial, o variables, cuando cambian en función de la actividad asistencial. Pueden subclasificarse en tangibles o intangibles en función de si se pueden medir con unidades monetarias o no.
 - Ejemplos de costes sanitarios tangibles: hospitalización del paciente, fármacos administrados, exploraciones complementarias, transporte sanitario, sueldo de los trabajadores sanitarios, hospitalización a domicilio...
 - Ejemplos de costes sanitarios intangibles: duelo, sufrimiento, ansiedad, cambio en el estilo de vida, adaptación a discapacidad...
- **Costes no sanitarios:**
 - **Directos.** Costes en los que incurre el paciente para recibir un tratamiento determinado. Pueden, a su vez, ser tangibles o intangibles. Ejemplos: precio del taxi para ir al hospital, tristeza por no asistir a una boda de un amigo por estar ingresado...
 - **Indirectos o de productividad.** Relacionados con la disminución de la productividad laboral. Ejemplos: días de baja laboral por enfermedad, disminución del rendimiento por secuelas de un ictus...
 - **Sociales.** Costes que sufragan algunas entidades públicas (servicios sociales) o personas cercanas al paciente (familiares, amigos...).

Otra clasificación más simple, aceptada por otros autores, consiste en la diferenciación de costes sanitarios directos, costes indirectos, costes no sanitarios y costes intangibles (Tabla 14.2).

TIPOS DE COSTES	EJEMPLOS
Sanitarios directos	Hospitalización del paciente, fármacos administrados, exploraciones complementarias, transporte sanitario, sueldo de los trabajadores sanitarios, hospitalización a domicilio...
Indirectos	Días de baja laboral por enfermedad, disminución del rendimiento por secuelas de un ictus...
No sanitarios	Precio del taxi para ir al hospital...
Intangibles	Duelo, sufrimiento, ansiedad, cambio en el estilo de vida, adaptación a discapacidad...

Tabla 14.2. Tipos de costes.



Recorda

- Los costes indirectos (no sanitarios) son aquellos relacionados con la disminución de la productividad laboral.

14.3. Evaluación de la eficiencia: análisis de costes

Las evaluaciones económicas de las tecnologías sanitarias tienen como objetivo evaluar la eficiencia de las intervenciones sanitarias. Pese a que estos análisis pueden realizarse desde la perspectiva del paciente o de un servicio de salud público o privado (en ese caso únicamente considerarían los costes directos), la mayoría de estos análisis se realizan bajo la perspectiva de la sociedad en global, teniendo en cuenta tanto los costes directos como los costes indirectos, dado que se dispone de un sistema sanitario público costeado a partir de impuestos de todos los contribuyentes **MIR 15-16, 195**. Además, dado que se basan en la población real, es más adecuado tener en cuenta los resultados en forma de efectividad, en lugar de eficacia.

Existen varios tipos de estudios o análisis, en función de si se tienen en cuenta resultados y/o costes, y si se comparan una o varias intervenciones (**Tabla 14.3**).

		¿SE EVALÚAN COSTES Y RESULTADOS?		
		No		Sí
		Solo resultados	Solo costes	
¿SE COMPARAN 2 O MÁS OPCIONES?	No	Estudio de consecuencia	Estudio de coste	Análisis coste-consecuencia
	Sí	Evaluación de eficacia o efectividad	Análisis de coste	A. coste-beneficio A. de minimización de costes A. coste-efectividad A. coste-utilidad

Tabla 14.3. Tipos de evaluaciones económicas.

De los anteriores, los más importantes son aquellos que comparan varias intervenciones teniendo en cuenta tanto los costes como los resultados. En todos ellos, los costes se expresan en unidades monetarias (€, \$, £, ¥...), mientras que las unidades de medida de los resultados dependen del tipo de análisis utilizado (**Tabla 14.4**) **MIR 18-19, 20; MIR 18-19, 124; MIR 16-17, 125; MIR 13-14, 208**.

TIPO DE ANÁLISIS	MEDIDA DE LOS RESULTADOS
Coste-beneficio	Unidades monetarias Euros, dólares, libras...
Coste-efectividad	Unidades naturales o físicas Años de vida ganados
Coste-utilidad	Unidades de utilidad (preferencias subjetivas) Años de vida ajustados por calidad (AVAC)
Minimización de costes	Unidades naturales o físicas Resultados similares entre intervenciones (fármacos genéricos)

Tabla 14.4. Medida de los resultados en los análisis de evaluación económica.

Recorda

- Los costes siempre se miden en unidades monetarias, pero los resultados dependen del tipo de análisis: unidades monetarias (coste-beneficio), años de vida ganados (coste-efectividad), años de vida ajustados por calidad (coste-utilidad).

En función del objetivo del análisis, es posible realizar:

- **Análisis coste-beneficio.** Al tener los resultados las mismas unidades que los costes, es el único que permite valorar económicamente una única alternativa respecto a no hacer nada (comparar si los gastos que comporta una intervención sanitaria son menores a los beneficios económicos que aportará a largo plazo). Sus limitaciones principales son la dificultad ética de cuantificar económicamente los resultados en salud y que no permite considerar los costes intangibles.
- **Análisis de minimización de costes.** Permite escoger la opción más económica entre varias que han demostrado, de forma estadísticamente significativa, igualdad de resultados. Es el análisis de coste utilizado para comparar fármacos genéricos con el fármaco de referencia **MIR 16-17, 126**.
- **Análisis coste-efectividad.** Es el análisis más utilizado y busca conocer la opción más eficiente, comparando el coste por cada unidad de efectividad (años de vida ganados, curación obtenida, casos evitados, incapacidad evitada, muerte evitada...). Existen análisis de coste-eficacia, cuando los resultados han sido obtenidos en condiciones ideales, pero no deberían usarse para las evaluaciones económicas poblacionales.
- **Análisis coste-utilidad.** Busca conocer la opción más eficiente, comparando el coste por cada unidad de utilidad (años de vida ajustados por calidad). Se puede considerar un subtipo de análisis de coste-efectividad.

Recorda

- Cuando dos alternativas presentan resultados idénticos, el análisis de minimización de costes permite conocer la alternativa más eficiente, por ser la más económica. Dicho análisis se utiliza para comparar genéricos con el fármaco de referencia.

Las unidades de utilidad tienen en cuenta tanto el beneficio objetivo (años de vida ganados) como las preferencias subjetivas de los pacientes, es decir, la calidad de vida relacionada con la salud (CVRS) medida mediante cuestionarios estandarizados con propiedades métricas demostradas **MIR 21-22, 49; MIR 19-20, 183; MIR 17-18, 217**.

Los años de vida ajustados por calidad (AVAC) o *Quality Adjusted Life Year* (QALY) corresponden a una ponderación de los años de vida ganados por una escala del 0 al 1 (donde 0 equivale al fallecimiento o al peor estado de salud posible y 1 al mejor estado de salud posible).

Los análisis de costes pueden tener un enfoque prospectivo, es decir, buscar la intervención más económica para obtener un determinado objetivo de salud; o retrospectivo, esto es, plantear nuevas alternativas más económicas a las ya existentes para mantener un mismo objetivo de salud.

Dado que los análisis de costes se realizan con los datos disponibles de una muestra determinada, al inferir los resultados en la población general existe el riesgo de cometer errores aleatorios debidos al muestreo (si la precisión es baja) y no aleatorios debidos a la existencia de sesgos (si la validez es baja). Este riesgo o incertidumbre en las evaluaciones económicas de intervenciones sanitarias se puede medir mediante ► **MIR 17-18, 216** :

- **Técnicas cualitativas o análisis de sensibilidad (AS).** Existen principalmente cuatro: AS univariante (el más utilizado), AS multivariante, AS de extremos (forma más básica de AS multivariante) y AS umbral.
- **Técnicas cuantitativas estadísticas.** Existen muchas: AS probabilístico (incluye las cadenas de Markov y la simulación de Montecarlo), intervalos de confianza del coste-efectividad incremental, cuasi-intervalos de confianza (*confidence box*), expansión de series de Taylor (método Delta), elipse de confianza (*confidence ellipse*), método de Fieller, método de muestreo repetitivo (*bootstrap*) no paramétrico.

■ Análisis incremental

El **análisis incremental** es de elección para el análisis y en la interpretación de los resultados en el análisis de coste-efectividad y el análisis de coste-utilidad. Consiste en comprobar si una alternativa es coste-efectiva o coste-útil respecto a otra en función de un umbral fijado *a priori* ► **MIR 14-15, 188** .

Para ello, inicialmente se ordenan las alternativas por orden creciente de efectividad, eliminando las que tengan un coste superior al de otra más efectiva (p. ej., si el fármaco A prolonga la esperanza de vida 5 años de media con un coste total de 5.000 €, mientras que el fármaco B prolonga la esperanza de vida 3 años de media con un coste total de 15.000 €; se escogería el fármaco A por ser más efectivo y más económico al mismo tiempo).

Posteriormente, se comparan aquellas opciones más efectivas pero menos económicas con aquellas otras menos efectivas pero más económicas, mediante el **índice o razón de coste-efectividad incremental (RCEI)**, utilizado en los análisis de coste-efectividad, que informa del coste por cada año de vida adicional ganado con la alternativa más eficaz y menos económica, y se escoge una u otra alternativa en función de si supera o no el umbral fijado inicialmente.

$$RCEI = \frac{\text{Coste1} - \text{Coste2}}{\text{Efectividad1} - \text{Efectividad2}}$$

Ejemplo 1. Supóngase que el fármaco A prolonga la esperanza de vida 5 años de media, con un coste total de 5.000 €, mientras que el fármaco B prolonga la esperanza de vida 6 años de media, con un coste total de 15.000 €. ¿Qué fármaco es más coste-efectivo para un umbral de coste-efectividad de 30.000 € por año de vida adicional ganado por paciente?

$$RCEI = \frac{5.000 - 15.000}{5 - 6} = \frac{-10.000}{-1} = 10.000 \text{ € por año adicional}$$

Tras calcular la RCEI, se escogería el fármaco B por presentar un coste-efectividad incremental 20.000 € por debajo del umbral de coste-efectividad.

Ejemplo 2. Supóngase que el fármaco A prolonga la esperanza de vida 5 años de media, con un coste total de 5.000 €, mientras que el fármaco B prolonga la esperanza de vida 6 años de media, con un coste total de 55.000 €. ¿Qué fármaco es más coste-efectivo para un umbral de coste-efectividad de 30.000 euros por año de vida adicional ganado por paciente?

$$RCEI = \frac{5.000 - 55.000}{5 - 6} = \frac{-50.000}{-1} = 50.000 \text{ € por año adicional}$$

Tras calcular la RCEI, se escogería el fármaco A, porque el fármaco B presenta un coste-efectividad incremental 20.000 € por encima del umbral de coste-efectividad.

En los análisis de coste-utilidad se emplea el **índice o razón de coste-utilidad incremental (RCUI)**. Conceptualmente es idéntico al RCEI, pero utilizando unidades de utilidad (AVAC) en lugar de las unidades físicas o naturales (años de vida ganados) usadas en los análisis de coste-efectividad.

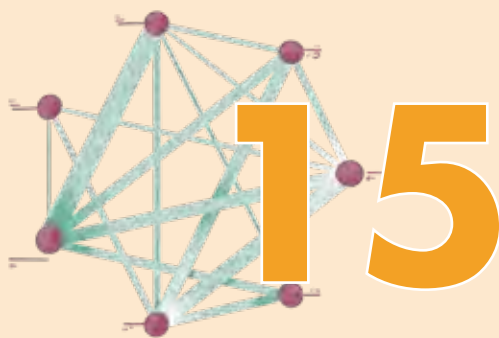
Otra forma, más grosera y menos ética, de comparar dos alternativas es mediante la **ratio coste-efectividad**, que consiste en ver qué alternativa es más eficiente teniendo en cuenta, únicamente, el coste por año. Por ejemplo, supóngase que el fármaco A prolonga la esperanza de vida 5 años de media, con un coste total de 5.000 €, mientras que el fármaco B prolonga la esperanza de vida 6 años de media, con un coste total de 15.000 €. La ratio coste efectividad del fármaco A es de 1.000 €/año y la del fármaco B es de 2.500 €/año. De esta manera se tiene una idea de qué fármaco es más eficiente por año, pero no del incremento del coste por cada año adicional del fármaco más efectivo, por lo que el análisis incremental es de elección frente a la ratio coste-efectividad simple.

Casos clínicos

En un ensayo clínico aleatorizado, enmascarado a doble ciego y controlado con el fármaco de referencia (atorvastatina 40 mg/dL) de un nuevo hipolipemiente (fármaco intervención), realizado en 1.200 pacientes. Al año del inicio del estudio, se obtienen los siguientes resultados: el grupo tratado con el fármaco intervención disminuye los niveles de LDL 63 mg/dL de media, mientras que en el grupo control (atorvastatina 40 mg/dL) disminuye los niveles de LDL 51 mg/dL de media ($p < 0,01$). Si el precio del nuevo fármaco fuera similar al de la atorvastatina:

- 1) El fármaco experimental es igual de eficaz y de eficiente que la atorvastatina.
- 2) El fármaco experimental es más eficaz e igual de eficiente que la atorvastatina.
- 3) El fármaco experimental es más eficaz y más eficiente que la atorvastatina.
- 4) El fármaco experimental es más efectivo y más eficiente que la atorvastatina.

RC: 3



La calidad asistencial y la seguridad del paciente

Orientación MIR

Segundo tema de Planificación y Gestión en interés; su rentabilidad ha crecido enormemente en las últimas convocatorias del MIR. Presta especial atención a las definiciones de calidad asistencial.

Preguntas MIR

- ▶ MIR 18-19, 126; MIR 18-19, 221
- ▶ MIR 15-16, 197
- ▶ MIR 13-14, 209

Conceptos clave

- La calidad asistencial comprende un conjunto de cualidades que integran tanto la calidad científico-técnica o intrínseca como la percibida o extrínseca.
- La calidad total supone un cambio de filosofía organizativa desde el modelo de burocracia profesional.
- Las guías clínicas, fruto de la medicina basada en la evidencia (MBE), deben tener más peso en la práctica que la experiencia profesional individual.
- El Modelo Europeo de Excelencia (EFQM) es un modelo industrial que permite a las organizaciones autoevaluarse y que fomenta la comunicación intrínseca con el objetivo de integrar iniciativas.
- La gestión por procesos incluye el conjunto de actividades que se interrelacionan mientras se presta un servicio: es típico de modelos de gestión modernos, multidisciplinarios y transversales como las unidades de gestión clínica.

15.1. Introducción

La **calidad asistencial**, de acuerdo con la OMS, es la característica administrativa que asegura que cada paciente reciba el conjunto de servicios diagnósticos y terapéuticos más adecuados para conseguir una asistencia óptima, teniendo en cuenta todos los factores y los conocimientos del paciente y del servicio médico, y lograr el mejor resultado con el mínimo riesgo de efectos iatrogénicos y la máxima satisfacción del trabajador y del paciente con el proceso ► MIR 13-14, 209.

Está, a su vez, compuesta por las tres definiciones del término:

- Calidad como el conjunto de cualidades de un producto satisfactorio.
- Calidad como función gestora que desarrolla ese conjunto de características o gestión de la calidad.
- Calidad como filosofía empresarial o calidad total.

15.2. Componentes de la calidad asistencial

La calidad asistencial como conjunto de cualidades incluye: efectividad y eficiencia, adecuación de la competencia profesional y científico-técnica, accesibilidad y disponibilidad, aceptabilidad, seguridad, satisfacción y continuidad. Las diferentes dimensiones de la calidad tienen relevancia relativa, según el actor del sistema sanitario, y se ha diferenciado la calidad percibida o extrínseca (que se relaciona con la disponibilidad, seguridad y satisfacción subjetiva del usuario) de la calidad científico-técnica o intrínseca (Tabla 15.1) ► MIR 15-16, 197.

ASPECTO DE LA CALIDAD	PACIENTES	SANITARIOS	GESTORES
Medios científico-técnicos	+	+++	-
Efectividad	++	+	-
Eficiencia	-	-	+++
Satisfacción	+++	+	+
Accesibilidad	+++	+	++

Tabla 15.1. Importancia relativa concedida por cada estamento a los diferentes atributos de la calidad asistencial.

La gestión de la calidad representa la función directiva que determina y aplica ese conjunto de cualidades mediante diferentes procesos: planificación, organización, control y mejora (asume que los modelos de gestión siempre son susceptibles de mejorar).

La calidad total es un conjunto de principios y metodología que busca la mejora continua por medio de la voluntad, la implicación, la ejemplaridad, la prevención de errores y la evaluación. Es un cambio en la filosofía organizativa frente al modelo previo de burocracia profesional (Tabla 15.2).

Recuerda

- El modelo de calidad total potencia la responsabilidad, mientras que el modelo de burocracia potencia la autonomía profesional.

CALIDAD TOTAL	BUROCRACIA PROFESIONAL
Responsabilidad colectiva	Responsabilidad profesional
Liderazgo de gestión	Liderazgo clínico
Responsabilidad	Autonomía
Evaluación en objetivo, proceso y resultado	Evaluación en resultado
Mejora continua	Control de calidad
Participación	Autoridad administrativa

Tabla 15.2. Modelo de calidad total frente a modelo de burocracia profesional.

15.3. Monitorización, evaluación, normalización y certificación de la calidad. Modelos de gestión de la calidad

- **Monitorizar** consiste en vigilar de manera continuada aspectos clave, como la incidencia de infección en herida quirúrgica o la tasa de complicaciones en el paciente diabético. Tiene el objetivo de detectar cambios o caídas de la calidad con rapidez.
- **Evaluar** la calidad asistencial supone comparar las estructuras, procesos y resultados con lo esperado o con el estándar, para resaltar los aspectos positivos e identificar áreas de fallo y oportunidades de mejora. Existen varias razones para evaluar la calidad asistencial: éticas, de eficiencia, de seguridad y de tipo científico-técnico.
- **Normalizar** es aportar soluciones para aplicaciones repetitivas, con el objetivo de optimizar el orden de una estructura o proceso para un determinado objetivo.
- **Certificar** es determinar, de manera independiente y por una entidad externa de reconocido prestigio, la conformidad de un producto, proceso o servicio respecto a una norma, reglamentaria o profesional.

En el ámbito sanitario, la búsqueda de la normalización de los procesos sanitarios ha permitido consolidar el gran avance científico. Una muestra es el número existente de guías clínicas y protocolos sanitarios, nacidos de la revisión sistemática de la literatura. Constituyen un pilar fundamental de la medicina basada en la evidencia (MBE) y son de mayor utilidad, desde el punto de vista de la gestión clínica, que la experiencia profesional individual.

Existen determinados **modelos de gestión de la calidad** que integran las funciones de normalización y certificación con el objetivo de la búsqueda de la mejora hacia la excelencia. Las diferencias entre los principales modelos, así como las características que orientan hacia cada uno de ellos ► MIR 18-19, 126, se resumen en la Tabla 15.3.

	EFQM (Modelo Europeo de Excelencia)	ISO (International Standards Office)	JCAHO (Joint Commission on Accreditation of Healthcare Organizations)
Método	Autoevaluación	Certificación externa	Acreditación externa
Ámbito	Toda la organización	Toda/parte de la organización	Toda la organización
Aplicabilidad	Cualquier organización	Cualquier organización	Específico sociosanitario
Limitaciones	Empresarial	Burocrática	Poco estratégica
Utilidad	Directivos	Mandos intermedios	Clínicos

Tabla 15.3. Principales modelos de gestión de la calidad.



El **Modelo Europeo de Excelencia (EFQM)** es un modelo industrial, creado por el empresariado europeo y avalado por la Unión Europea, que permite a las organizaciones autoevaluarse y fomenta la comunicación intrínseca con el objetivo de integrar iniciativas. Es el estándar de excelencia en el momento actual. Consta de nueve criterios, entre los que se incluye la gestión por procesos.

En cada uno de ellos, para obtener un certificado, la organización tendrá que someterse a una auditoría por la entidad certificadora.

Recuerda

- El EFQM es el estándar de excelencia que se está imponiendo en nuestro medio. Incluye la gestión por procesos: una organización del trabajo multidisciplinar.

Clásicamente, los sistemas sanitarios se han estructurado siguiendo una **gestión por organigrama**. Se trata del esquema de servicios más empleado en nuestros hospitales (Neurología, Cardiología, etc.). Cada vez más, esta estructura se está demostrando menos eficaz en un contexto de alta complejidad como es el medio sanitario.

Se denomina **gestión por procesos** a una concepción de los flujos de trabajo multidisciplinar y multidepartamental. Se entiende como proceso el conjunto de actividades mutuamente relacionadas o que interactúan mientras se presta un servicio. Busca generar estructuras flexibles, autónomas, transversales y que se adapten a escenarios cambiantes, como, por ejemplo, las unidades de gestión clínica. El EFQM es un ejemplo de gestión por procesos.

15.4. La seguridad del paciente

En el año 1999, en Estados Unidos se publicó el célebre informe *To Err is Human: Building a Safer Health System*, que situó los eventos adversos relacionados con la asistencia sanitaria como la séptima causa de muerte en todo el país (entre 44.000 y 98.000 fallecimientos anuales). En la actualidad, nadie debería dudar que se está ante una actividad que implica multitud de riesgos. Se denomina **seguridad del paciente** al conjunto de estrategias que buscan la reducción del riesgo de daño innecesario asociado a la atención sanitaria hasta un mínimo aceptable. Este mínimo está en función de los conocimientos del momento, los recursos disponibles y el contexto en el que se presta la atención, ponderados frente al riesgo de no dispensar ese tratamiento/cuidado o de dispensar otro.

En España, la **Estrategia de Seguridad del Paciente del Sistema Nacional de Salud (período 2015-2020)** recoge un conjunto de acciones con el objetivo de promover y mejorar la cultura de la seguridad en las organizaciones sanitarias, incorporar la gestión del riesgo sanitario, formar a los profesionales y a los pacientes en aspectos básicos de seguridad del paciente e implementar prácticas seguras, implicando a pacientes y ciudadanos.

No todos los fallos que se producen en los centros sanitarios tienen las mismas repercusiones sobre los pacientes. Se habla de **incidente relacionado con la seguridad del paciente** ante cualquier evento o circunstancia que ha ocasionado o podría haber ocasionado un daño innecesario a un paciente.

Dentro de estos incidentes, un **cuasincidente** será aquel que no alcanza al paciente, mientras que un **incidente sin daño** alcanza al paciente, pero no causa ningún daño apreciable. Obviamente, la peor de las situaciones será el **evento adverso**: incidente que produce daño a un paciente.

En el ámbito organizativo, existen dos maneras o modelos de gestionar estas desagradables situaciones (Tabla 15.4).

MODELO CENTRADO EN LAS PERSONAS	MODELO CENTRADO EN LOS SISTEMAS/ORGANIZACIONES
Los actos son atribuibles plena y exclusivamente a cada individuo	Los errores son consecuencia del sistema
Solución: culpabilizar y castigar (<i>name, blame & shame</i>)	Solución: analizarlos y aprender

Tabla 15.4. Modelos de gestión del riesgo (Reason J. Human error: models and management. *BMJ*. 2000;320:768-770).

Obviamente, en el contexto de la medicina del siglo XXI, debe prevalecer el **modelo centrado en los sistemas/organizaciones**, donde lo más importante cuando surge un problema no es quién, sino qué, cómo, cuándo, dónde, por qué y cómo se podría haber evitado. El objetivo no es cambiar la condición humana, sino los factores sistémicos.

Ejemplo. Un médico residente, a las 3:00 a.m. en una de sus primeras guardias en el Servicio de Urgencias, prescribe amoxicilina a una paciente alérgica a dicho antibiótico. El modelo centrado en las personas señalaría: ¿cómo no ha mirado antes el residente que la paciente era alérgica?; en cambio, el modelo relativo a los sistemas y organizaciones ampliaría las preguntas: ¿existía en ese momento un adjunto supervisando?, ¿el residente había descansado las horas suficientes?, ¿los sistemas de alarma de la prescripción electrónica no han impedido llegar a esa situación? Como se ve, un análisis más profundo y constructivo.

Reason, uno de los grandes teóricos de la seguridad del paciente, explicaba esta situación con la analogía a través de un queso suizo **MIR 18-19, 221**. Cada una de las lonchas del queso sería una barrera que impide que se produzca un evento adverso: supervisión de un adjunto, descanso adecuado, alarmas en el sistema, etc.

Cuando se produce un agujero en cada una de las lonchas, en cada una de las barreras, y además se producen en el mismo momento (están alineados), es cuando uno se encuentra ante un evento adverso (Figura 15.1).

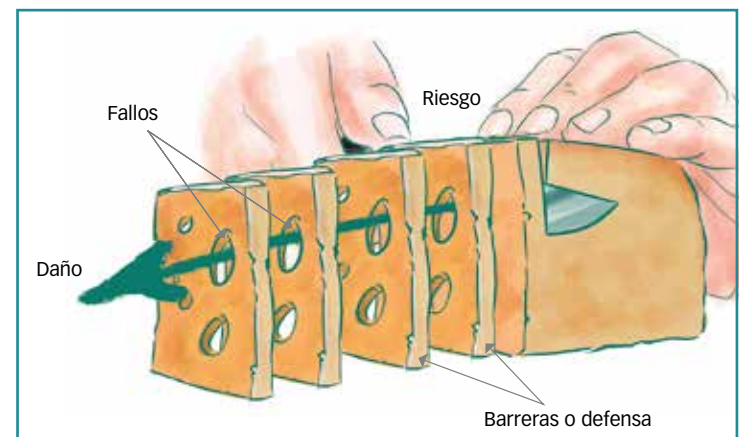


Figura 15.1. Modelo del queso suizo en la producción de accidentes. (Reason J. Human Error: models and management. *BMJ*; 2000;320:768-770).

Para terminar, las comunidades autónomas poseen **sistemas de notificación de incidentes** específicos que, de una manera anónima, voluntaria y confidencial, permiten mejorar la seguridad de los pacientes a partir del análisis de situaciones concretas que produjeron, o podrían

haber producido, daño a los pacientes. El **Plan de Calidad para el Sistema Nacional de Salud** ha desarrollado una plataforma común, el Sistema de Notificación y Aprendizaje para la Seguridad del Paciente (SiNASP).



Bibliografía

- ⊙ Aranaz JM, Aibar C, Vitaller J, Mira JJ (dirs). Gestión sanitaria: calidad y seguridad de los pacientes. Madrid. Díaz de Santos, 2014.
- ⊙ Dawson B, Trapp R. Bioestadística médica, 4.ª ed. Manual Moderno, 2005.
- ⊙ Donabedian A. La calidad de la atención médica. Ediciones científicas. La Prensa Médica Mexicana, S. A. México DF, 1984.
- ⊙ Greenberg RS. Epidemiología médica, 4.ª ed. Manual Moderno, 2005.
- ⊙ Hernández Aguado I, et al. Manual de Epidemiología y Salud Pública, 2.ª ed. Editorial Médica Panamericana, 2011.
- ⊙ Informe Plan de Calidad para el Sistema Nacional de Salud 2006-2010. Ministerio de Sanidad y Política Social, 2009.
- ⊙ Jiménez Jiménez J (coord.). Manual de Gestión para jefes de Servicios Clínicos, 2.ª ed. Díaz de Santos, 2000.
- ⊙ Martínez-González MA, Sánchez Villegas A, Faulin Fajardo FJ. Bioestadística amigable, 3.ª ed. Elsevier, 2014.
- ⊙ Piédrola Gil G, et al. Medicina Preventiva y Salud Pública, 11.ª ed. Elsevier-Masson, 2008.
- ⊙ Ribera J, Gutiérrez Fuentes JA, Rosenmöller M. Gestión en el sector de la Salud. Elementos de Gestión en las Instituciones. Pearson Educación S. A., 2006.
- ⊙ Ruiz M, Morillo LE. Epidemiología clínica: investigación clínica aplicada. Editorial Médica Panamericana, 2004.
- ⊙ San Martín H, Pastor V. Salud comunitaria: teoría y práctica. Díaz de Santos, 1984.
- ⊙ Simón P, Hernando P, Martínez S, et al. Estándares éticos para una organización sanitaria excelente. *Rev Calid Asist.* 2003;18(6): 379-404.
- ⊙ Varo J. Gestión estratégica de la calidad en los servicios sanitarios. Un modelo de gestión hospitalaria. Díaz de Santos, 1994.

En la realización de este manual han intervenido:

Coordinación del proyecto multidisciplinar:

Sandra León Paraíso, Fernando de Teresa Galván y Fernando de Teresa Mansilla

Coordinación editorial:

Sonia Clavería Irazo

Corrección:

Fernando Sáez Domingo y Equipo CTO

Ilustración de cubierta:

Txema Martínez Ávila

Ilustración de interior:

Marina Heredia Marcos

Diseño de interiores:

Txema Martínez Ávila y Victoria Rivas Guerra

Algoritmos y otros gráficos:

Victoria Rivas Guerra

Maquetación e impresión:

Cañizares Artes Gráficas

Material digital:

Adaptación digital:

Óscar Díaz González, Adrián Gutiérrez Álvarez, Juan Carlos Quintero de Teresa y José Antonio Serrano Sánchez

Equipo multimedia:

Cecilia León Paraíso, María López Romero, Daniel Paraíso Frías, Beatriz Parrondo Jiménez y Miriam Rodríguez Paredes

Agradecimientos:

Alba Bernad Sanz, Roxana Cerdá Cosme, Magdalena Espigares Rodríguez, Sergio Mata Redondo, Laura Miño Cifuentes y Laura Ponce Pastor